STOCHASTIC APPROXIMATION OF THE MULTIDIMENSIONAL GENERALIZED GRADED UNFOLDING MODEL

A Dissertation
Presented to
The Academic Faculty

by

David R. King

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Psychology

Georgia Institute of Technology
August, 2017

STOCHASTIC APPROXIMATION OF THE MULTIDIMENSIONAL GENERALIZED GRADED UNFOLDING MODEL

Approved by:

Dr. James S. Roberts, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Christopher Hertzog
School of Psychology
Georgia Institute of Technology

Dr. Susan E. Embretson
School of Psychology
Georgia Institute of Technology

Dr. Daniel H. Spieler
School of Psychology
Georgia Institute of Technology

Dr. Brian Habing
Department of Statistics
University of South Carolina

Date Approved: May 01, 2017

ACKNOWLEDGEMENTS

First and foremost, I want to thank my doctoral advisor and committee chair, Jim Roberts, for his guidance and support throughout this project. I am deeply appreciative of his mentorship, and the quality of the final paper is a reflection of his close collaboration on the project.

Additionally, I am grateful for the assistance of Taylor Baldwin and Eliana Dubin, who spent a considerable amount of time helping me run (and troubleshoot) the simulations for this study.

I am thankful to my committee members Susan Embretson, Brian Habing, Christopher Hertzog, and Daniel Spieler for offering a number of helpful suggestions throughout the proposal and defense stages.

Finally, I could not have completed this project without the love and support of my wife, Jamie King.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The multidimensional generalized graded unfolding model (MGGUM; Roberts & Shim, 2010) is a distance-based, unfolding multidimensional item response theory (MIRT) model for measuring person and item characteristics from graded or binary disagree-agree responses to Thurstone or Likert style questionnaire items. It can also be used when graded satisfaction or preference responses to more general multidimensional stimuli (e.g., photographs of faces, samples of coffee, etc.) are obtained. The item parameters in the MGGUM have previously been estimated using Markov chain Monte Carlo (MCMC; Roberts & Shim, 2010) and marginal maximum a posteriori (MMAP; Thompson, 2014) procedures, although neither procedure efficiently estimates higher-dimensional MGGUMs.

An efficient estimation procedure for the MGGUM will increase the utility of the model for applied researchers. One candidate procedure is the Metropolis-Hastings Robbins-Monro (MH-RM; Cai, 2010a; Cai, 2010b; Cai, 2010c) algorithm, which has been shown to efficiently estimate other multidimensional models. Considering the efficiency of the MH-RM algorithm for estimating MIRT models, the current paper examined the utility of the MH-RM algorithm for estimating item parameters in the MGGUM.

Initial attempts to estimate the MGGUM with the MH-RM resulted in severe misestimation of item parameters, although estimation accuracy was markedly improved through modifications to the MH-RM. Namely, the Newton-Raphson step for updating item parameters was replaced with the L-BFGS-B method for constrained optimization (Byrd, Lu, Nocedal, & Zhu, 1995). The improved estimation accuracy was likely the result

of using a method that did not require inversion of a Hessian matrix, and that updated the item parameter estimates multiple times on each stage of the algorithm.

For the modified MH-RM algorithm to be a useful method for estimating the MGGUM, the algorithm should be computationally fast, and the accuracy of the estimates should be comparable to alternative estimation methods. Therefore, the current study examined both the speed of the algorithm and the accuracy of the estimates through a parameter recovery study that varied test length (10, 20, or 30 items), sample size (1000, 1500, or 2000 persons), number of response categories (2, 4, or 6), dimensional structure of items (simple or complex), and dimensionality (2 or 3 dimensions). Further, the MMAP procedure was used to obtain item parameter estimates on a subsample of the design to facilitate direct comparison between the modified MH-RM and MMAP results.

Results from the parameter recovery study indicated that the modified MH-RM was faster than the MMAP method for three-dimensional models, and that differences in estimation accuracy between the modified MH-RM and MMAP methods were small. The runtime patterns observed for two- and three-dimensional models indicated that increasing the number of dimensions in the model led to small, linear increases in runtime for the modified MH-RM, and relatively large, exponential increases in runtime for the MMAP method. For the modified MH-RM method, a three-dimensional model was estimated in approximately 68.73 minutes, a 24% increase in runtime over estimation of a two-dimensional model (55.32 minutes). For the MMAP method, a three-dimensional model was estimated in approximately 220.68 minutes, an 807% increase in runtime over estimation of a two-dimensional model (27.35 minutes). Although it is important to note

that the MMAP method completed in approximately half the time as the modified MH-RM method for the two-dimensional model.

Examination of parameter recovery indicated that item discrimination and location parameters were estimated more accurately with the MMAP method by a small margin over the modified MH-RM method, although subjective response category threshold and person parameters were estimated at essentially the same level of accuracy for both methods. Increasing the number of response categories in the model led to improved estimation of item discrimination parameters and person parameters, with larger improvements observed for the increase from 2 to 4 response categories than for the increase from 4 to 6 response categories. Complex item structure led to improved estimation over simple structure of subjective response category threshold parameters. However, complex item structure led to decreased estimation accuracy of item location and person parameters. As expected, increasing the test length led to improved estimation of person parameters, with larger improvements observed for the increase from 10 to 20 items than for the increase from 20 to 30 items.

A real data analysis examined the utility of the modified MH-RM algorithm for estimating MGGUM parameters in practical situations. The analysis was performed on 1237 graded disagree-agree responses to 24 emotion stimuli, and uncovered an interpretable, three-dimensional solution that corresponded well with the circumplex theory of emotion (Russell, 1980).

The current study contributed to the development of the MGGUM and to the feasibility of using the model for applied measurement problems. The modified MH-RM

and MMAP methods were implemented within R package *mirt*, and will be released to the general public in the near future. This will give applied researchers a convenient means for conducting multidimensional analyses on preference, satisfaction, or self-similarity data.

INTRODUCTION

The multidimensional generalized graded unfolding model (MGGUM; Roberts & Shim, 2010) is a distance-based, unfolding multidimensional item response theory (MIRT) model for measuring person and item characteristics from graded or binary disagree-agree responses to Thurstone or Likert style questionnaire items. It can also be used for measurement in cases where respondents indicate their level or preference, satisfaction, or self-similarity with more general stimuli (e.g., preference for consumer products, satisfaction with services, self-similarity to emotion faces, etc.). For simplicity, these more general stimuli will be referred to as "items" in this proposal and they will be characterized with "item parameters" from the model.

The item parameters in the MGGUM have been estimated previously using Markov chain Monte Carlo (MCMC; Roberts & Shim, 2010) and marginal maximum a posteriori (MMAP; Thompson, 2014) procedures, although neither procedure efficiently estimates higher-dimensional MGGUMs (i.e., 3 or more dimensions). MCMC is a sampling-based, fully Bayesian estimation method in which a joint posterior distribution of all model parameters is constructed. For a highly parameterized model, such as the MGGUM, the MCMC method can take days to complete depending on the number of iterations required to to achieve convergence of the chain and the computer software that is used to implement the technique. MMAP is a maximum-likelihood based technique in which prior distributions are specified for all model parameters and person parameters, θ_j, are integrated out of the likelihood function via numerical quadrature. Quadrature methods typically rely on a rectangular grid of quadrature points, which are used for numerical

integration. This method is efficient for lower-dimensional MGGUMs, although it becomes inefficient as the number of dimensions in the model increases, because higher dimensional models require the evaluation of an exponentially higher number of quadrature points.

An efficient estimation procedure for the MGGUM will increase the practicality of the model for applied researchers. One candidate procedure is the Metropolis-Hastings Robbins-Monro (MH-RM; Cai, 2010a; Cai, 2010b; Cai, 2010c) algorithm, which has been shown to efficiently estimate parameters in other multidimensional models including IRT models. The current study reviews the literature related to the MGGUM and the MH-RM algorithm, and then examines the utility of a modified version of the algorithm for estimating parameters in the MGGUM.

1.1 Background

1.1.1 Unfolding Models

The MGGUM is classified as an *unfolding* model because it assumes that the expected value of a response increases as the distance between the person and item location in *D*-dimensional space decreases. This expected value function is nonmonotonic and single-peaked. Unfolding models have been shown to fit certain types of questionnaire data, such as attitude (Roberts, Laughlin, & Wedell, 1999), personality (Stark, Chernyshenko, Drasgow, & Williams, 2006), and vocational interest (Tay, Drasgow, Rounds, & Williams, 2009) ratings better than *dominance* models that assume the expected value of a response increases as the latent score increases.

The idea that the expected value of a response is maximized when the distance between the latent score and item location is zero can be traced back to Thurstone's (1928) scoring method for attitude measurement. Thurstone calculated a score by taking the median scale value of all endorsed statements. This method was attractive for its simplicity, although limited because it did not take into account levels of endorsement (i.e., graded responses), and did not directly parameterize the relationship between the latent score and the item location.[1] Both of these limitations were overcome by unfolding item response theory (IRT) models.

Davison (1977) presented an unfolding model for continuous responses that produced nonmonotonic and single-peaked expected value functions.[2] Namely, he modeled the *j*th person's response to the *i*th item, z_{ij}, by squaring the difference between the locations of the *j*th person, θ_j, and the *i*th item, δ_i:

$$z_{ij} = a_j(\theta_j - \delta_i)^2 + b_j \tag{1}$$

The squared term ignores the sign of the difference, $\theta_j - \delta_i$, effectively making $(\theta_j - \delta_i)^2$ a measure of the squared *distance* between the *j*th person and the *i*th item. Andrich (1988; 1989) used this idea to develop the squared simple logistic model, the first parametric unfolding IRT model for binary responses.

1.1.2 The Generalized Graded Unfolding Model

[1] However, Thurstone's diagnostic for identifying relevant items implicitly invoked the notion of a proximity-based response process

[2] Note that the notation used in this paper is different from the notation used in Davison (1977).

Roberts (Roberts, 1995; Roberts & Laughlin, 1996; Roberts, Donoghue & Laughlin, 2000) developed several unfolding IRT models for binary or graded responses by modeling an *observed* response as a function of two *subjective* responses. Andrich and Luo (1993) had previously used this idea to develop the hyperbolic cosine model (HCM), an unfolding IRT model for binary responses. [3] To distinguish between an observed and subjective response, consider an observed rating scale, from *strongly disagree* = 0 to *strongly agree* = *C*. The *j*th person's observed response to the *i*th item, z_{ij}, contains information about the distance of the *j*th person from the *i*th item, but it does not contain information about whether the person is located below or above the item on the latent scale. To indicate whether the person disagreed or agreed with the statement from below or above the item, the rating scale would need to include twice as many response categories to account for all possible levels of disagreement and agreement from below and above the item.

Suppose that an unobserved, subjective rating scale, from *strongly disagree from below* = 0 to *strongly disagree from above* = *M*, accounts for all levels of disagreement and agreement, both below and above the item. (Note that $M = 2C + 1$ and *strongly agree from below*/*strongly agree from above* are the two middle categories in the subjective rating scale). Let the subjective response *from below* for the *j*th person on the *i*th item be denoted as y_{ij1}, and the subjective response *from above* be denoted as y_{ij2}. Each observed response is mapped onto two corresponding subjective responses, $\{z_{ij}\} \rightarrow \{y_{ij1}, y_{ij2}\}$, such that $\{0\} \rightarrow \{0, M\}, \{1\} \rightarrow \{1, M-1\}, ..., \{C\} \rightarrow \{C, M-C\}$. Namely, each observed response is

[3] The relationship between observed and subjective responses is slightly different in Andrich's unfolding IRT models than in Roberts' unfolding IRT models. Namely, in Andrich's models, the observed response representing the strongest level of agreement corresponds with a *single* subjective response. In Roberts' models, the observed response representing the strongest level of agreement corresponds with *two* subjective responses.

mapped to one subjective response from below the midpoint of the subjective rating scale and to one subjective response from above the midpoint. An illustration of the relationship between the observed and subjective rating scales is shown in Figure 1.

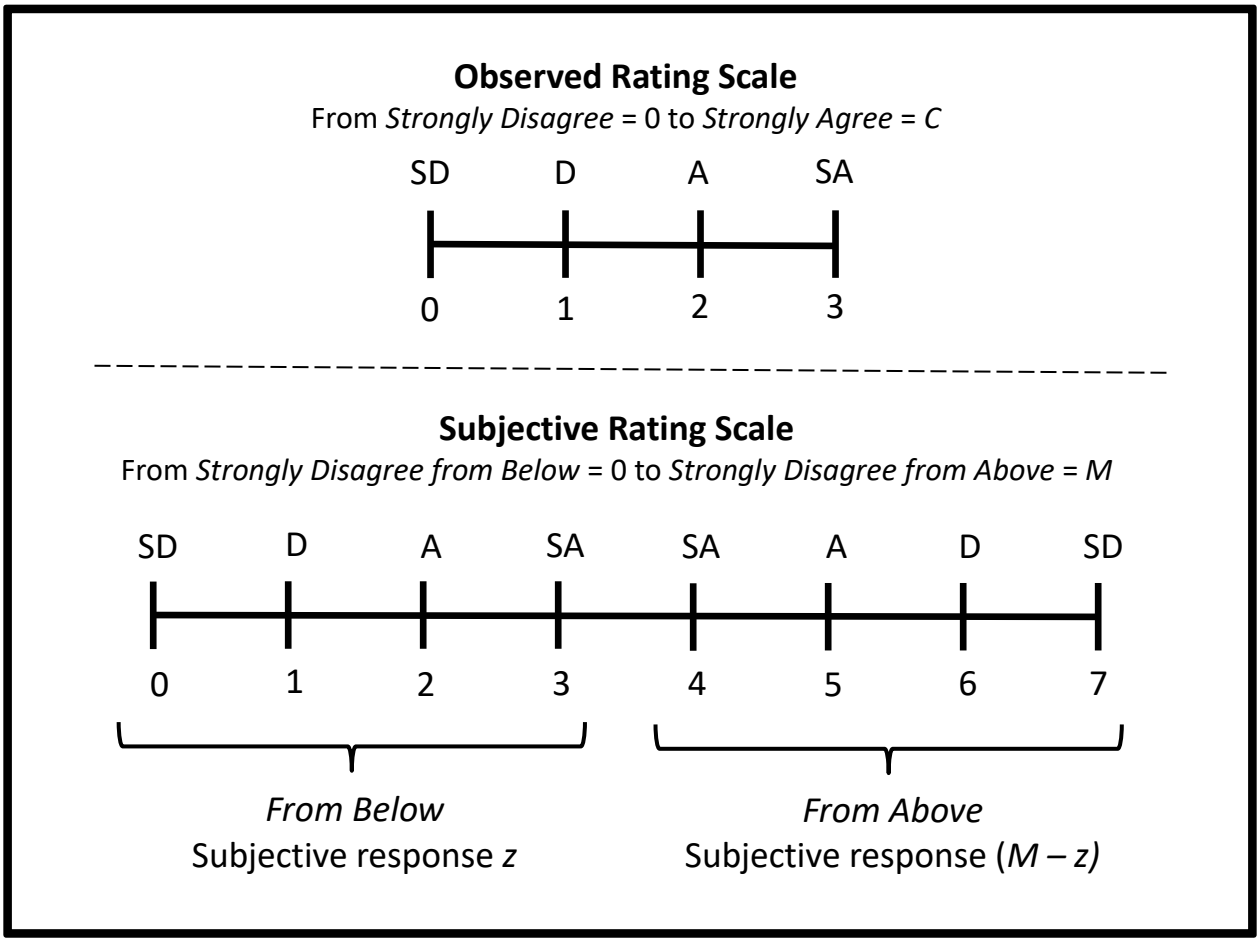

Figure 1 - Illustration of the observed and subjective rating scales for the generalized graded unfolding model with four observed response categories. Note. SD = Strongly Disagree; D = Disagree; A = Agree; SA = Strongly Agree.

Assuming that the subjective responses are monotonically related to $\theta_j - \delta_i$, the probability function for the *j*th person's subjective response to the *i*th item is:

$$P\left[Y_{ij} = y \middle| \theta_j\right] = \frac{\exp\left(\alpha_i\left[y(\theta_j - \delta_i) + \sum_{k=0}^{y}\tau_{ik}\right]\right)}{\sum_{w=0}^{M}\left[\exp\left(\alpha_i\left[w(\theta_j - \delta_i) + \sum_{k=0}^{w}\tau_{ik}\right]\right)\right]}, \quad (2)$$

given the constraint that

$$\sum_{k=0}^{M}\tau_{ik} = 0, \quad (3)$$

where α_i is the discrimination of the *i*th item, τ_{ik} is the relative location of the *k*th subjective response category threshold of the *i*th item, and $\tau_{i0} = 0$ by definition. Summing the probabilities for the subjective response from below, $P\left[Y_{ij1} = z \middle| \theta_j\right]$, and the subjective response from above, $P\left[Y_{ij2} = M - z \middle| \theta_j\right]$, gives the probability of an observed response in the generalized graded unfolding model (GGUM; Roberts, Donoghue & Laughlin, 2000).[4]

$$P\left[Z_{ij} = z \middle| \theta_j\right] = \frac{\exp\left(\alpha_i\left[z(\theta_j - \delta_i) + \sum_{k=0}^{z}\tau_{ik}\right]\right) + \exp\left(\alpha_i\left[(M - z)(\theta_j - \delta_i) + \sum_{k=0}^{z}\tau_{ik}\right]\right)}{\sum_{w=0}^{C}\left[\exp\left(\alpha_i\left[w(\theta_j - \delta_i) + \sum_{k=0}^{w}\tau_{ik}\right]\right) + \exp\left(\alpha_i\left[(M - w)(\theta_j - \delta_i) + \sum_{k=0}^{w}\tau_{ik}\right]\right)\right]}. \quad (4)$$

The model assumes that thresholds are symmetric about the point, $\theta_j - \delta_i = 0$, which leads to the following two definitions:

[4] This formulation of the GGUM is slightly modified from the GGUM given in Roberts, Donoghue, and Laughlin (2000). Namely, in the current formulation the summed threshold term, , is added, rather than subtracted, toand . Estimated thresholds will therefore have the opposite sign (i.e., positive instead of negative). This formulation was used to allow for the specification of lognormal priors for the estimated threshold parameters implemented in this study.

$$\tau_{ik} = -\tau_{i(M-k+1)} \text{ for } k \neq 0, \quad (5)$$

and

$$\tau_{i(C+1)} = 0. \quad (6)$$

The GGUM is a unidimensional, divide-by-total, unfolding IRT model. The model is unidimensional in that it measures a single latent score for each person and is divide-by-total in that the denominator is simply the sum of the numerators.

As noted by Roberts et al. (2000), the shape of the expected value function in the GGUM is symmetric about the point $(\theta_j - \delta_i) = 0$. Additionally, it is determined by both the discrimination and threshold parameters. Holding the thresholds constant while increasing the item discrimination will increase the effect of the distance between θ_j and δ_i on the expected value of the response. Namely, as the item discrimination approaches infinity, the distance between θ_j and δ_i will completely determine the expected value of the response in a deterministic fashion. In contrast, increasing the interthreshold distance while holding the item discrimination constant has the opposite effect on the shape of the expected value function. Namely, the function becomes less steep as the thresholds move farther apart.

1.1.3 Estimation Procedures for the GGUM

Several estimation procedures have been used to obtain estimates for GGUM parameters. Roberts (1995; & Laughlin, 1996) initially used a joint maximum likelihood

(JML; Birnbaum, 1968) procedure to estimate both item and person parameters for a constrained version of the GGUM (i.e., the graded unfolding model or GUM). This method was limited because the JML estimates did not converge to the true values as the sample size increased (i.e., the estimates were inconsistent), and the maximum likelihood algorithm often became stuck at local maxima.

The item parameters ($\delta_i, \alpha_i, \tau_{ik}$) in the unconstrained GGUM were estimated (Roberts et al., 2000) by a marginal maximum likelihood (MML; Bock & Aitkin, 1981) approach in which the person parameters, θ_j, are integrated out of the likelihood function via an expectation-maximization (EM) algorithm. The removal of theta from the likelihood function helped to eliminate the problems of inconsistent estimates and local maxima observed with the JML procedure. (The local maxima were not eliminated with MML, although they were not a problem given that the starting values were in the neighborhood of the solution.) The MML estimates of item parameters in conjunction with the item responses were used to estimate θ_j parameters via an expected a posteriori (EAP; Bock & Mislevy, 1982) procedure. This procedure is Bayesian in the sense that a prior distribution is specified for theta and a posterior distribution is computed for each person. The mean of the posterior distribution is the EAP estimate and the standard deviation of the posterior distribution is the standard error of the estimate.

A fully Bayesian, MCMC approach was implemented by De la Torre, Stark, and Chernyshenko (2006) for estimating both person and item parameters in the GGUM. In this approach, prior distributions are specified for all model parameters. A Markov chain is then constructed from a Metropolis-Hastings sampler (Metropolis, Rosenbluth,

Rosenbluth, Teller, & Teller, 1953; Hastings, 1970) and the parameter estimates are determined from the means of the joint posterior distribution. The MCMC procedure produced estimates and standard errors that were generally more accurate than their MML counterparts, as indicated by lower root mean squared errors.

More recently, Roberts and Thompson (2011) used a marginal maximum a posteriori (MMAP; Mislevy, 1986) approach for estimating item parameters in the GGUM. In the MMAP procedure, a prior distribution is specified for each item parameter, and the likelihood function is weighted by the point densities of these prior distributions to produce a posterior distribution. MMAP is similar to MML in that person parameters are integrated out of the likelihood function. In Roberts and Thompson (2011), the MMAP procedure produced parameter estimates that were generally more accurate than the estimates produced by MCMC or MML procedures. The largest differences in accuracy were observed when the number of response categories was small (i.e., 2 or 4). In this situation, both the MMAP and MCMC procedures produced more accurate estimates than the MML procedure. Because the MMAP procedure was more efficient than the MCMC procedure, MMAP was recommended as the best estimation procedure for the GGUM.

1.1.4 The Multidimensional Generalized Graded Unfolding Model

MIRT models are required for instances when stimuli represent more than one latent dimension. Recent attempts have been made to extend unidimensional, unfolding IRT models to the multidimensional case. For example, Javaras and Ripley (2007) gave the multidimensional unfolding model (MUM), a between-item multidimensional model for graded response data. Wang and Wu (2015) also gave a between-item multidimensional

unfolding model, the confirmatory multidimensional generalized graded unfolding model (CMGGUM), which is a multidimensional extension of the GGUM for confirmatory analyses. However, multidimensional items do not always adhere to *simple* structure, a necessary condition for the specification of a confirmatory, between-item multidimensional model. For an assessment to adhere to simple structure, the items must form relatively homogenous, unidimensional clusters. Namely, items within a cluster should measure the same dimension and items between clusters should measure different dimensions.

In some instances, items cannot be neatly partitioned into homogenous clusters. In these instances, the items are said to display *complex* structure. Moreover, there are many instances in which the underlying dimensional structure of the data is not well understood, and therefore, the degree to which each items discriminates among latent dimensions cannot be specified prior to estimation. In these exploratory instances, the MGGUM provides a flexible framework for estimating items that display both between-item and/or within-item dimensionality.

The MGGUM is a multidimensional extension of the GGUM.[5] The response function for the MGGUM is defined as

$$P\left[Z_i = z \middle| \underline{\theta}_j\right] = \frac{\exp\left[\left(z\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right)+\sum_{k=0}^{z}\psi_{ik}\right]+\exp\left[\left((M-z)\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right)+\sum_{k=0}^{z}\psi_{ik}\right]}{\sum_{w=0}^{C}\left(\exp\left[\left(w\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right)+\sum_{k=0}^{w}\psi_{ik}\right]+\exp\left[\left((M-w)\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right)+\sum_{k=0}^{w}\psi_{ik}\right]\right)} \quad (7)$$

[5] The GGUM is a special case of the MGGUM when the number of dimensions is equal to unity.

where D is the number of dimensions, $\underline{\theta_j}$ is a vector of latent scores for person j, θ_{j1} , ..., θ_{jD}; $\underline{\delta_i}$ is a vector of location coordinates for the ith item, δ_{i1} , ..., δ_{iD}; $\underline{\alpha_i}$ is a diagonal matrix of discrimination parameters, α_{i1} , ..., α_{iD}; and ψ_{ik} is the k^{th} multidimensional subjective response category threshold for the i^{th} item. These multidimensional thresholds are defined as

$$\psi_{ik} = \sum_{d=1}^{D} \alpha_{id} \tau_{ik} \tag{8}$$

where the τ_{ik} (and hence, ψ_{ik}) are assumed to be constant across the D dimensions of the multidimensional space. Additionally, τ_{i0} (and therefore ψ_{i0}) is equal to zero by definition.

The MGGUM is noncompensatory in the sense that the expected value of a response decreases as the distance between the person and item increases. Namely, the person must be close to the item on all dimensions to endorse the item, and closeness on one dimension does not compensate for lack of closeness on another dimension. However, it should be noted that noncompensatory models are often distinguished from compensatory models in the literature (e.g., Reckase, 2009) by the mathematical form of the response function. Namely, compensatory models sum a function of the model parameters across dimensions, whereas noncompensatory models multiply this function across dimensions. Although the MGGUM response function does not multiply response probabilities across dimensions, it is similar to other noncompensatory MIRT models, such as Sympson's (1978) multidimensional three-parameter logistic model and Whitely's (1980) multicomponent latent trait model, in that a location parameter is estimated on each dimension that an item measures.

Another defining feature of the MGGUM is that it cannot be rotated (Roberts & Shim, 2010). Namely, there is an optimal orientation for the solution that is determined during the estimation procedure. This is similar to Carroll and Chang's (1970) INDSCAL model in the multidimensional scaling literature, and also to Sympson's (1978) noncompensatory MIRT model. The MGGUM cannot be rotated because the discrimination matrix for each item is diagonal, and rotating the solution would typically lead to a non-diagonal matrix, changing the definition of the model. Considering this feature, rotational constraints are not specified prior to estimation.

1.1.5 Estimation Procedures for the MGGUM

MGGUM solutions have been obtained using MCMC (Roberts & Shim, 2010) and MMAP (Thompson, 2014) procedures. Thompson (2014) compared these procedures for a two-dimensional MGGUM in a parameter recovery study. She found estimation accuracy to be similar between the MCMC and MMAP solutions, although the MMAP procedure was considerably more efficient than the MCMC procedure. Namely, the MMAP procedure finished, on average, in 7 minutes, and the MCMC procedure finished in 5,029 minutes.

In addition to increased efficiency, the MMAP procedure was preferred over the MCMC procedure because the MCMC procedure suffered from a number of technical and theoretical problems. Namely, local traps were occasionally encountered, the meaning of dimensions sometimes switched during estimation, and some of the most extreme theta values switched signs. It was unclear whether these problems resulted from uninformative starting values or the MCMC procedure itself.

Although the MMAP procedure is viable for two-dimensional solutions, the procedure will become inefficient as the number of dimensions increases because the efficiency of the procedure is dependent on the number of quadrature points evaluated. Assuming that rectangular quadrature is used for computing the marginal posterior likelihood, then the number of quadrature points evaluated in the MMAP procedure increases exponentially as the number of dimensions in the MGGUM increases linearly. Namely, the total number of quadrature points evaluated across all dimensions is $T = Q^D$, where Q is the number of quadrature points evaluated for a single dimension and D is the number of dimensions.

For example, suppose 30 quadrature points were evaluated for each dimension. This would result in 900 quadrature points for a two-dimensional model, 27,000 points for a three-dimensional model, and 810,000 points for a four-dimensional model. This pattern of increasing inefficiency as the number of dimensions increases is often referred to as the *curse of dimensionality* (Bellman, 1957), and suggests that the MMAP procedure will become prohibitively inefficient for higher-dimensional solutions.

One way to mitigate the inefficiency problem is through adaptive quadrature (Schilling & Bock, 2005), which reduces the number of quadrature points per dimension. Or to avoid the problem altogether, a stochastic approximation method could be used such as Monte Carlo quadrature with Gibbs sampling (Meng & Schilling, 1996). This method is similar to the MH-RM algorithm, although Monte Carlo integration does not take into account the random variation introduced into the theta estimates from the Gibbs sampler.

1.1.6 The Metropolis-Hastings Robbins-Monro Algorithm

In response to the inefficiency of quadrature-based methods for estimating high-dimensional MIRT models, Cai (2010a, 2010b, 2010c) proposed the MH-RM algorithm. The algorithm combines the Metropolis-Hastings (MH; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953; Hastings, 1970) sampler used in Bayesian estimation methods with the Robbins-Monro (RM; Robbins & Monro, 1951) root-finding algorithm for noise-corrupted data.

The MH-RM algorithm is similar to JML estimation in that the MH-RM algorithm samples latent scores by conditioning on provisional item parameter estimates, and then updates item parameter estimates by conditioning on provisional latent scores. However, Cai (2010b) mentions that the JML procedure will not necessarily converge because it does not properly account for the uncertainty in the latent scores. Namely, the tentative latent scores in the JML procedure are treated as fixed effects. The MH-RM algorithm overcomes this problem by using the MH sampler to introduce uncertainty into the latent scores. Further, the MH-RM method overcomes the issue of incidental parameters in JML by integrating theta out of the likelihood function. This is done by sampling theta from its posterior distribution.

1.1.7 Details of the MH-RM Algorithm

The MH-RM algorithm is an iterative procedure, where each iteration is composed of three steps: stochastic imputation, stochastic approximation, and the Robbins-Monro update. A single iteration of the algorithm is described in detail below. Let * denote a value(s) that has been updated on the *t*th iteration of the algorithm. Further, let *Z* denote

the item responses, Θ denote the latent scores, Ψ denote the item parameters, and $X = (Z, \Theta)$.[6]

1.1.7.1 Stochastic Imputation

A Metropolis-within-Gibbs procedure (Chib & Greenberg, 1995) is used to sample *m* sets of missing data, $\Theta_s^*(\text{where } s = 1, 2, ..., m)$, from transition kernel, $f(\Theta_s \mid Z, \Psi)$ to form *m* sets of complete data, $X_s^* = (Z, \Theta_s^*)$. Namely, draws 1 to *N* from the MH sampler on a given iteration of the algorithm are determined by the sequence:

1. Draw $\theta_1^* \sim \xi(\theta_1 \mid \theta_2, ..., \theta_N, Z, \Psi)$,
2. Draw $\theta_2^* \sim \xi(\theta_2 \mid \theta_1^*, \theta_3 ..., \theta_N, Z, \Psi)$,
...
N. Draw $\theta_N^* \sim \xi(\theta_N \mid \theta_1^*, ..., \theta_{N-1}^*, Z, \Psi)$,

where $\xi(\theta_j \mid \theta_1, ..., \theta_N, Z, \Psi)$ denotes the full conditional density for θ_j.

1.1.7.2 Stochastic Approximation

Noise-corrupted gradient vectors and Hessian matrices are calculated for each item by taking the first and second partial derivatives, respectively, of the complete data log likelihood function with respect to each item parameter. These item-specific gradient vectors and Hessian matrices are combined, respectively, into a single gradient vector,

[6] The notation used in the current paper is different than the notation used in Cai (2010b). In this paper, item responses and latent scores are denoted with Z and , respectively, for consistency with the IRT notation used earlier in the paper. In Cai (2010b), responses and latent scores are denoted with Y and X respectively. This latter notation is helpful when considering IRT models within a generalized linear modeling framework.

$\nabla(\Psi \mid X_s^*)$, and block diagonal Hessian matrix, $\mathrm{H}(\Psi \mid X_s^*)$. The Hessian matrix, $\mathrm{H}(\Psi \mid X_s^*)$, is then used to compute a recursive approximation of the Fisher Information matrix:

$$\Gamma^* = \Gamma + \gamma_t \left(\frac{1}{m} \sum_{s=1}^{m} \mathbf{H}(\Psi \mid X_s^*) - \Gamma \right), \tag{9}$$

where γ_t is a Robbins-Monro gain constant. (γ_t starts at unity for iteration t = 1 and decreases to zero as $t \to \infty$).

1.1.7.3 Robbins-Monro Update

The item parameters, Ψ, are recursively updated according to the equation:

$$\Psi^* = \Psi + \gamma_t \left(\left[\Gamma^* \right]^{-1} \frac{1}{m} \sum_{s=1}^{m} \nabla(\Psi \mid X_s^*) \right). \tag{10}$$

1.1.8 Efficiency of the MH-RM Algorithm

To examine the efficiency of the MH-RM algorithm, Cai (2010b) estimated the parameters in a five-dimensional, exploratory, item factor analytic model with 120 item parameters using the MH-RM algorithm and Bock and Aitkin's (1981) EM algorithm (i.e., MML). He fit the model to a dataset with 753 persons and 24 items, each with five graded response categories. The EM algorithm used rectangular quadrature with five points per dimension, for a total of 5^5 = 3,125 quadrature points. The MH-RM algorithm finished in 95 seconds, whereas the EM algorithm finished in 1 hour and 27 minutes.

Cai (2010b) noted that the relationship between the number of dimensions in the model and the estimation efficiency of the MH-RM algorithm is linear, rather than exponential. This indicates that the MH-RM algorithm does not suffer from the curse of dimensionality that plagues the MML and MMAP procedures.

1.2 Objectives of the Current Study

Considering the efficiency of the MH-RM algorithm for estimating MIRT models, the current paper examined the utility of the MH-RM algorithm for estimating the item parameters in the MGGUM. For the MH-RM algorithm to be a useful method for estimating the MGGUM, the algorithm should be efficient, and the accuracy of the estimates should be comparable to alternative estimation methods. Therefore, the current study examined both the efficiency of the algorithm and the accuracy of the estimates in comparison to the MMAP procedure. To facilitate comparison between the MH-RM and MMAP procedures, the MH-RM algorithm included prior distributions on all item parameters. Data demands, estimation accuracy, and runtime efficiency for the MH-RM procedure were assessed through a parameter recovery study. Furthermore, the practical utility of the procedure was explored through a real data analysis.

CHAPTER 2. METHOD

2.1 Modifications to the MH-RM Algorithm

Initial attempts to estimate the MGGUM resulted in severe misestimation of item parameters. It was hypothesized that misestimation occurred on account of two characteristics of the MH-RM algorithm. The first is that the algorithm performs only a single Newton-Raphsen update of item parameters. In contrast, other estimation methods, including MMAP, update item parameters iteratively until a specified convergence criteria is met. The second characteristic hypothesized to result in misestimation is that the MH-RM requires inversion of the Hessian matrix, and the Hessian was observed to be non-invertible on some iterations of the algorithm.[7]

The limited-memory BFGS method with boundary constraints (L-BFGS-B; Byrd, Lu, Nocedal, & Zhu, 1995) resolves both of the aforementioned limitations of the Newton-Raphson step in the MH-RM. It performs iterative updating of item parameters until a specified convergence criteria is met, and it does not require inversion of a Hessian matrix. The method directly estimates the inverse of the Hessian, rather than calculating the Hessian and then solving for its inverse. The algorithm ensures that the Hessian is positive definite, which in turn ensures that the approximated inverse exists. This approach makes the method robust to situations where the Hessian cannot be inverted, and improves the efficiency of the algorithm because taking the inverse of a large matrix is a computationally expensive procedure. Additional efficiency gains result from the *limited-memory* approach. Namely, the method stores a relatively small number

[7] In the event that the Hessian could not be inverted, the generalized inverse was taken instead.

of values that can be used to reproduce the (approximated) inverse Hessian, rather than storing the full inverse Hessian. The inverse Hessian is then updated across cycles, rather than recalculated for each cycle. Finally, the L-BFGS-B method is attractive because of its allowance for boundary constraints to be imposed on the solution. This ensures that alpha and threshold parameter estimates in the MGGUM are nonnegative, and it prevents against Heywood cases, which have been observed in other MGGUM estimation methods (Thompson, 2014).

The L-BFGS-B method can be implemented within the MH-RM algorithm without decreasing the utility of the MH-RM for high-dimensional item response theory models. The method does not suffer from increases in runtime as the number of dimensions increases, in contrast to quadrature-based approaches. In the current study, the method was implemented in place of the Netwon-Raphsen step for updating the item parameters in Equation 10. (Note that the Robbins-Monro gain constant in Equation 10 is still applied to the item parameter update values.)

2.2 Parameter Recovery Simulation

2.2.1 Experimental Design

A parameter recovery study was built from a factorial design that included five factors: test length (10, 20, or 30 items), sample size (1000, 1500, or 2000 persons), number of response categories (2, 4, or 6), dimensional structure of items (simple or complex), and dimensionality (2 or 3 dimensions). Thirty replications were conducted for each cell of the design for a total of 3 *x* 3 *x* 3 *x* 2 *x* 2 *x* 30 = 3,240 observations. The levels were selected to facilitate comparison between the results in the current study and results from previous

MGGUM estimation studies using MMAP (Thompson, 2014) and MCMC (Roberts & Shim, 2010) methods.

The modified MH-RM algorithm was used for model estimation in all cells of the design. Moreover, the MMAP procedure was also used for model estimation in a subset of the replications for the purpose of directly comparing the estimation accuracy and computational efficiency of the modified MH-RM and MMAP procedures. This subset included five effects: estimation method (modified MH-RM or MMAP), test length (10 or 30 items), sample size (1000 or 2000 persons), dimensional structure (simple or complex) and dimensionality (2 or 3 dimensions). Five replications were randomly sampled from each cell in this design subset, for a total of $2 \times 2 \times 2 \times 2 \times 5 = 80$ replications in which model estimates were obtained using both the modified MH-RM and MMAP procedures. The number of response categories in the model was held constant at four.

In comparing the relative efficiency of the modified MH-RM and MMAP methods, it is important to note that the runtimes of the methods are greatly influenced by their implementations. Namely, the programming language, structure of the code, and other implementation details substantially affect the speed of the methods. The current study attempted to minimize implementation differences between the MH-RM and MMAP methods by using the same code for the maximization step of each algorithm (i.e., the L-BFGS-B method for updating item parameter estimates). Further, the same programming languages were used for each method. Namely, the algorithms were implemented in R, and numerical functions such as partial derivatives and the complete data log likelihood were implemented in C/C++.

2.2.2 Measures of Speed and Estimation Accuracy

The primary focus of the parameter recovery study was to examine the speed and estimation accuracy of the modified MH-RM. Speed was measured by runtime (in seconds) of the estimation program, and depended on operating system, processor speed, random-access memory (RAM), and other software and/or hardware specifications. Therefore, runtime was measured with respect to a given system. Measurements of runtime were taken across multiple systems with different specifications. These measurements were converted to the scale of a reference system through the following transformation:

$$\mathrm{RT}_{\mathrm{ref}} = b_0 + b_1 \mathrm{RT}_{alt}, \tag{11}$$

where $\mathrm{RT}_{\mathrm{ref}}$ is the runtime on the reference scale from a preselected computer, RT_{alt}, is the runtime on an alternative computer, and b_0 and b_1 are the intercept and slope, respectively, for linearly transforming RT_{alt} onto the scale of $\mathrm{RT}_{\mathrm{ref}}$. The b_0 and b_1 parameters were estimated by regressing $\mathrm{RT}_{\mathrm{ref}}$ onto RT_{alt}. Seven data points for each measure were used to compute the regression coefficients for each linear transformation. The data points were obtained by calculating runtime on the reference computer and the alternative computer using seven common test datasets.

Estimation accuracy was measured with root-mean-square deviation (RMSD), which is an index of the average discrepancy between estimated and true values. Consider an arbitrary type of parameter from the MGGUM (i.e., θ, α, δ, or τ) and denote this given parameter type as β. On a given replication, the RMSD for parameter β is equal to:

$$\text{RMSD} = \left\{ \frac{\sum_{n=1}^{N} \left[\hat{\beta}_n - \beta_n \right]^2}{N} \right\}^{1/2} \quad (12)$$

where

$\hat{\beta}_n$ is the estimated value of the *n*th parameter;

β_n is the true value of the *n*th parameter; and

N is the total number of these parameters on a given replication (e.g., 20 item locations).

For simple item structure conditions, only item locations on measured dimensions were included in the calculation of the item location *RMSD*. For complex structure item conditions, all item locations were included in the calculation of the item location *RMSD*.

2.2.3 Data Generation

2.2.3.1 True Parameter Values

Person parameters, θ_{jd}, were sampled from a multivariate normal distribution with uncorrelated dimensions, $\underline{\Theta}_j \sim MVN(\underline{0}, I)$ where $\underline{0}$ represents a null vector and I represents an identity matrix. Item parameters ($\delta_{id}, \alpha_{id}, \tau_{ik}$) were constructed from unidimensional GGUM item parameter estimates derived from an MMAP procedure in a previous study (Roberts & Thompson, 2011).

The unidimensional items were partitioned into five adjacent intervals across the theta continuum. In constructing the multidimensional items, an equal number of unidimensional items were sampled with replacement from each theta interval to ensure that there would be adequate information across the multidimensional latent space. After sampling, the unidimensional items were randomly sorted to ensure that each item had an equal probability of representing each dimension.

The specific method for constructing multidimensional items depended on the item loading structure (simple or complex) and the number of dimensions (two or three). In the simple structure (i.e., between-item multidimensionality) conditions , each multidimensional item was constructed from a single unidimensional item. Namely, for the two-dimensional conditions, sampled discriminations and locations were assigned to the first dimension for odd numbered items and to the second dimension for even numbered items. Thresholds were assigned to be equivalent to their unidimensional counterparts. (Note that thresholds in the MGGUM are dimensionless and therefore did not need to be assigned to a particular dimension.) Moreover, the discriminations and locations were fixed to zero on the second dimension for odd-numbered items and to zero on the first dimension for even-numbered items. For the three-dimensional conditions, sampled discriminations and locations for every third item (i.e., 1, 4, …) were assigned to the first dimension, sampled discriminations and locations for items 2, 5, …, were assigned to the second dimension, and sampled discriminations and locations for items 3, 6, …, were assigned to the third dimension. Discriminations and locations were fixed to zero on all other dimensions.

In the complex structure conditions, multidimensional items (i.e., within-item multidimensionality) were constructed by sampling with replacement one unidimensional item for each dimension. Namely, two unidimensional items were used for constructing each multidimensional item in the two-dimensional conditions, and three unidimensional items were used for constructing each multidimensional item in the three-dimensional conditions. The discrimination and location of the first unidimensional item was assigned as the discrimination and location on the first dimension (for 2D & 3D conditions).

Similarly, the discrimination and location of the second undimensional item was assigned to the second dimension (for 2D & 3D conditions), and the discrimination and location of the third unidimensional item was assigned to the third dimension (for only 3D conditions). Finally, thresholds were derived by averaging the *D* sets of unidimensional thresholds.

To keep the item information comparable between the items in the simple structure and complex structure conditions, the discriminations in the complex structure conditions were rescaled by $\frac{1}{\sqrt{D}}$. Initial testing of this approach indicated that the rescaling of the discriminations had the unexpected consequence of reducing the marginal response probability of the lowest response category (i.e., *strongly disagree* for graded responses & *disagree* for binary responses). Namely, the discriminations were reduced to the extent that it became unlikely for a response from the lowest response category to be generated.

A pilot study found that the marginal response probability for the first response category was 0.078 for the two-dimensional complex structure conditions, and 0.017 for the three-dimensional complex structure conditions.[8] The low probabilities led to instances when the first response category was never used for an item.

It was hypothesized that this problem could be resolved by rescaling the response category thresholds (in addition to rescaling the item discriminations). An algorithm was developed in which the marginal response probabilities for simple structure items were compared to the marginal response probabilities for complex structure items (constructed from the same parameters as the simple structure items). For each iteration of the algorithm, the complex structure item thresholds were rescaled by $1/L$, where $L = 1.0$ on the first iteration, and increased by 0.01 on each subsequent iteration until $L = 2.5$ was reached. The

[8] Models with four response categories were used in the pilot study.

absolute difference between the simple structure response probabilities and the complex structure response probabilities was then recorded for each iteration. Following the iterative procedure, the rescaling factor that minimized the absolute difference between the simple structure response probabilities and the complex structure response probabilities was selected. The rescaling factor was approximately 1.41 for the two-dimensional condition and approximately 1.73 for the three-dimensional condition. (Note that these values are approximately equivalent to $\sqrt{D}$.) Therefore, for the complex structure conditions, the thresholds were rescaled by multiplying the original thresholds by $\frac{1}{\sqrt{D}}$. These rescaling factors produced comparable marginal response category probabilities between simple and complex item structure conditions.

2.2.3.2 Response Generation

The true theta and item parameter values were used to calculate the probability of a response in a given category, z_{ij} = 0 to C_i, The observed response was obtained by sampling from a multinomial distribution with corresponding probability values. If a simulated person *j* used the lowest response category (i.e., response = 0) for all item responses, the person was removed from the data set.[9]

2.2.4 Parameter Estimation

2.2.4.1 Starting Values

[9] Persons not providing at least one response > 0 cannot be located on the underlying dimensions.

Starting values for the item parameters in the MGGUM were derived using a similar procedure as was used in Thompson (2014). Namely, starting values for item locations, $\tilde{\delta}_{id}$, were obtained by performing a detrended correspondence analysis (DCA; Hill & Gauch, 1980) on the item responses and taking the first *D* columns of item scores as the starting values. Starting values for discriminations, $\tilde{\alpha}_{id}$, were set to unity on all dimensions.

Starting values for thresholds, $\tilde{\tau}_{ik}$, were obtained using a prediction equation that was developed from a regression analysis on previous unidimensional MMAP estimates.[10,11] Namely, the initial values for thresholds $\tilde{\tau}_{i1}$ to $\tilde{\tau}_{iC}$ (note that $\tilde{\tau}_{i0}$ is fixed to zero) were a function of an origin constant, O_i, and an interthreshold distance constant, Δ_i:

$$O_i = 1.002 + 0.449\sqrt{\sum_{d=1}^{D}\tilde{\delta}_{id}^2} - 0.093V_i, \text{ and} \tag{13}$$

$$\Delta_i = 0.921 + 0.058\sqrt{\sum_{d=1}^{D}\tilde{\delta}_{id}^2} - 0.129V_i, \tag{14}$$

where V_i indicates the number of response categories for the *i*th item. The start value for the *k*th threshold of the *i*th item was calculated from the following equation:

$$\tilde{\tau}_{ik} = O_i + \Delta_i(C_i - k_i). \tag{15}$$

[10] Note that the parameterization of the MGGUM in this paper gives thresholds with positive signs. These reflected thresholds are opposite in sign (i.e., in comparison to the thresholds in earlier parameterizations of the GGUM).

[11] This same approach was taken in earlier GGUM research (e.g., Roberts & Thompson, 2011), although the prediction equation in the earlier research was developed from MML estimates instead.

Note that the starting value for the *C*th threshold, $\tilde{\tau}_{iC}$ (i.e., the threshold closest to the point, $\theta_j - \delta_i = 0$), was simply the origin, O_i. The starting value for each subsequent threshold was calculated by adding the interthreshold distance, Δ_i, to the starting value for the previous threshold. As the number of response categories, V_i, increased, the interthreshold distance, Δ_i decreased. Also note that the extremity of the thresholds depended on the extremity of the associated location parameters. If a given starting value, $\tilde{\tau}_{ik}$, was less than or equal to the arbitrary value of 0.1, then it was set to 0.1 to ensure that the starting value was positive and nonzero. (The 0.1 cutoff has been used in previous MGGUM analyses, e.g., Thompson, 2014.)

2.2.4.2 Prior Distributions

Prior distributions were specified for each parameter in the MGGUM. A multivariate normal prior was specified for the latent scores, Θ, with a mean vector (i.e., centroid) of zeros and covariance matrix equal to *I*, where *I* denotes an identity matrix.

For the item parameter prior distributions, a normal prior was specified for item locations, $\delta_{id} \sim N(0, 4)$ and a censored normal prior was specified for item discriminations: $\alpha_{id} \sim N(0,1)$, which naturally imposed the constraint that $\alpha_{id} \geq 0$. A lognormal prior was specified for thresholds with $\tau_{ik} \sim \ln N(\mu_{\tau_{ik}},1)$, where $\mu_{\tau_{ik}}$ was derived from the associated starting value for threshold $\tilde{\tau}_{ik}$:

$$\mu_{\tau_{ik}} = \ln \tilde{\tau}_{ik} - 1/2 , \tag{16}$$

which follows from the fact that $\tilde{\tau}_{ik} = e^{\left(\mu_{\tau_{ik}} + 1/2\right)}$. This approach for specifying threshold prior distributions was previously used with the MMAP procedure (Roberts & Thompson, 2011) for estimating the unidimensional GGUM.

2.2.4.3 Convergence Criteria

The MH-RM and MMAP solutions were considered to have converged when the maximum change in the item parameter estimates from iteration *t* to iteration *t*+1 was smaller than .001. The maximum number of iterations was set at 2000 iterations, although in the current study all replications converged before reaching this number.

2.2.4.4 Burn-in Period and Robbins-Monro Gain Constant

Cai (2010b) noted that the MH-RM algorithm worked well when the Robbins-Monro gain constant, γ_t = 1 for the first *B* iterations of the algorithm, where *B* represents the number of iterations in the burn-in period. Following the burn-in period, γ_t is a decreasing function that is applied to the item parameter update step to ensure that the convergence criteria is eventually met. Initial testing of the algorithm suggested that *B* = 500 resulted in relatively stable parameter estimates. This value was used for all modified MH-RM replications in the full design.

For the Robbins-Monro gain constant, Cai (2010b) noted that any decreasing function for updating γ_t across iterations of the MH-RM algorithm could be used given that it adheres to the following constraints:

$$\gamma_t \in (0,1], \sum_{t=1}^{\infty} \gamma_t = \infty, \text{ and } \sum_{t=1}^{\infty} \gamma_t^2 < \infty . \quad (17)$$

The current study used $\gamma_t = (.1/t)^{.75}$ as the decreasing function because the function had been used in previous MH-RM analyses (e.g., Chalmers & Flora, 2014), and it appeared to work well during initial testing of the algorithm.

2.2.4.5 Item Parameter Estimates and Associated Standard Errors

Following convergence of the modified MH-RM algorithm, item parameter estimates were obtained from $\Psi^{(t+1)}$, where *t* in this instance denotes the final iteration of the algorithm. Standard errors were derived from the approximate Fisher Information matrix. (Recall that Γ_{t+1} is computed during the second stage of the MH-RM algorithm). Cai (2008b) showed that Γ_{t+1} converged to the observed Fisher Information matrix as $t \rightarrow \infty$.

Standard errors were obtained by taking the inverse of Γ_{t+1} (i.e., Γ_{t+1}^{-1}), where Γ_{t+1} was calculated from Equation 9 and *m* was set to 5 (i.e., the number of calculated Hessian matrices). Γ_{t+1}^{-1} is the asymptotic variance-covariance matrix, and taking the square roots of the diagonal elements gives the standard errors for the corresponding parameter estimates.

2.2.4.6 Person Parameter Estimates and Associated Standard Errors

Person parameters were estimated using the EAP (Bock & Mislevy, 1982; Muraki & Carlson, 1995) procedure described in the introduction. Namely, 21 quadrature points were used for each dimension for a total of $21^2 = 441$ points for the 2D conditions and 21^3

= 9261 points for the 3D conditions. The minimum and maximum quadrature points were specified as –4 and +4 respectively, and the remaining quadrature points were equally spaced across the latent space. The mean, $\hat{\theta}_j$, and standard deviation, $\text{S.E.}_{\hat{\theta}_j}$, of the posterior distribution for the *j*th person were approximated with the following equations:

$$\hat{\theta}_{jd} \approx \frac{\sum_{q_D=1}^{Q_D} \cdots \sum_{q_2=1}^{Q_2} \sum_{q_1=1}^{Q_1} A_{qd} L\left(\mathbf{X}_j \middle| A_{q1}, A_{q2}, \ldots, A_{qD}\right) W\left(A_{q1}\right) W\left(A_{q2}\right) \ldots W\left(A_{qD}\right)}{\sum_{q_D=1}^{Q_D} \cdots \sum_{q_2=1}^{Q_2} \sum_{q_1=1}^{Q_1} L\left(\mathbf{X}_j \middle| A_{q1}, A_{q2}, \ldots, A_{qD}\right) W\left(A_{q1}\right) W\left(A_{q2}\right) \ldots W\left(A_{qD}\right)}, \text{and} \quad (18)$$

$$\text{S.E.}_{\hat{\theta}_{jd}} \approx \left[\frac{\sum_{q_D=1}^{Q_D} \cdots \sum_{q_2=1}^{Q_2} \sum_{q_1=1}^{Q_1} \left(A_{qd} - \hat{\theta}_{jd}\right)^2 L\left(\mathbf{X}_j \middle| A_{q1}, A_{q2}, \ldots, A_{qD}\right) W\left(A_{q1}\right) W\left(A_{q2}\right) \ldots W\left(A_{qD}\right)}{\sum_{q_D=1}^{Q_D} \cdots \sum_{q_2=1}^{Q_2} \sum_{q_1=1}^{Q_1} L\left(\mathbf{X}_j \middle| A_{q1}, A_{q2}, \ldots, A_{qD}\right) W\left(A_{q1}\right) W\left(A_{q2}\right) \ldots W\left(A_{qD}\right)} \right]^{1/2}, \quad (19)$$

where A_{qd} is one of Q quadrature points on the *d*th dimension, $W\left(A_{qd}\right)$ is the point density of the prior distribution associated with quadrature point A_{qd}, and $L\left(\mathbf{X}_j \middle| A_{q1}, A_{q2}, \ldots, A_{qD}\right)$ is the conditional likelihood of the *j*th individual's response vector given that the individual is located at the *q*th quadrature point on the *d*th dimension, A_{qd}.

2.2.4.7 Details of the MMAP Procedure

Item prior distributions and start values for the MMAP procedure were identical to those used for the MH-RM procedure. For the expectation step, rectangular quadrature was used with twenty-one quadrature points specified for each dimension. The maximization step used the L-BFGS-B method for constrained optimization, the same method used in the modified MH-RM. Item parameter estimates were obtained from the item parameter estimates on the final iteration before convergence. Item parameter standard errors were

obtained using the same approach as previously described for the modified MH-RM method. Namely, an approximated Fisher Information matrix was obtained through stochastic imputation of theta, by conditioning on item parameter estimates from the MMAP solution. The standard errors were then computed by taking the inverse of the Fisher Information matrix and calculating the square roots of the elements on the diagonal.

2.2.4.8 Rescaling Parameter Estimates and Associated Standard Errors

Parameter estimates and associated standard errors were rescaled to truth prior to the calculation of the RMSD statistics. This was necessary because the prior distributions placed on the item parameters overconstrained the solution.

An intercept-slope transformation equation was obtained for each dimension by regressing the true theta values for a given dimension on the corresponding estimated theta values. The intercept and slope were then used to place the theta estimates and location estimates on the same scale as their true counterparts. The threshold parameter estimates were rescaled by multiplying the estimates by the slope, and the discrimination parameter estimates were rescaled by multiplying the estimates by the reciprocal of the slope.

Standard errors were rescaled in a similar fashion. Namely, the standard errors for the theta estimates, location estimates, and threshold estimates were rescaled by multiplying the estimates by the slope. The standard errors for the discrimination estimates were rescaled by multiplying the estimates by the reciprocal of the slope.

2.2.5 Software

R scripts were developed for constructing items, generating latent scores and item responses, calculating model start values, and calculating parameter recovery statistics.

Additionally, several R packages were used in the analysis. R package *vegan* (Oksanen et al., 2007) was used for obtaining the DCA start values for item locations, package *car* (Fox & Weisberg, 2011)) was used for the ANOVA analysis and computation of Type III sums of squares, and package *ggplot2* (Wickham, 2009) was used for creating the figures in this paper.

R package *mirt* (Chalmers, 2012) was modified by this student to accommodate estimation of the MGGUM via the modified MH-RM and MMAP algorithms. The open-source nature of the R statistical software allowed for direct implementation of the MGGUM response function, new item parameter type definitions, and first and second partial derivatives of the complete data log likelihood with respect to each parameter. The new code was added to the package source files, which were then compiled and used to build an executable file for the updated package.

CHAPTER 3. RESULTS

3.1 Parameter Recovery Simulation

Five separate between-subjects factorial ANOVAs were conducted to examine the effect of test length, sample size, number of response categories, number of dimensions, and item structure on runtime and estimation accuracy. Namely, one univariate ANOVA was conducted for each of the four parameter types in the MGGUM ($\alpha, \beta, \tau,$ and θ) and for the runtime (RT) measure.

The five-way ANOVA models examined all main effects and interactions for the independent variables. This included five main effects, ten two-way interactions, ten three-way interactions, five four-way interactions and one five-way interaction. The Type I error rate was specified at $\alpha = 0.05/5 = 0.01$ because the same ANOVA model was used to examine each of the dependent variables ($RMSD_{\alpha}$, $RMSD_{\delta}$, $RMSD_{\tau}$, $RMSD_{\theta}$, & RT) in separate univariate analyses. Additionally, interpretation of effects was limited to those that were both statistically significant and associated with an eta-squared value of 0.10 or more.[12] Effects with eta-squared values of this magnitude were deemed to be reasonably strong and worthy of interpretation. All effects with eta-squared values of 0.10 or more were also statistically significant, essentially making effect size the only criterion for interpretation. Interpretable main effects were followed by post hoc paired comparisons and interpretable interaction effects were followed by post hoc interaction contrasts. The results from the analysis are shown in Tables 1 and 2.

[12] Eta-squared values were calculated from Type III sums of squares, although the sums of squares type selected for this analysis was inconsequential because of the balanced factorial design.

Table 1 - Average RMSD of parameter estimates and average runtime of estimation method by condition

	RMSD				
Effect	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\tau}$	$\hat{\theta}$	RT
Sample Size					
1000	0.285	0.381	0.338	0.431	2,534
1500	0.252	0.340	0.308	0.429	3,717
2000	0.230	0.320	0.292	0.427	4,947
Test Length					
10	0.313	0.430	0.312	0.556	1,975
20	0.233	0.316	0.306	0.402	3,799
30	0.220	0.296	0.319	0.329	5,424
Response Categories					
2	0.375	0.391	0.259	0.526	1,852
4	0.220	0.316	0.313	0.389	3,594
6	0.173	0.334	0.366	0.372	5,752

Note. $\hat{\alpha}$ = item discrimination estimate; $\hat{\delta}$ = item location estimate; $\hat{\tau}$ = subjective response category threshold estimate; $\hat{\theta}$ = person parameter estimate; RT = runtime (in seconds).

Table 1 (Continued) - Average RMSD of parameter estimates and average runtime of estimation method by condition

	RMSD				
Effect	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\tau}$	$\hat{\theta}$	RT
Item Structure					
Complex	0.227	0.457	0.172	0.485	3,689
Simple	0.284	0.237	0.453	0.373	3,777
Dimensionality					
2	0.229	0.285	0.289	0.370	3,054
3	0.282	0.409	0.337	0.488	4,411
Overall	0.256	0.347	0.313	0.429	4,081

Note. $\hat{\alpha}$ = item discrimination estimate; $\hat{\delta}$ = item location estimate; $\hat{\tau}$ = subjective response category threshold estimate; $\hat{\theta}$ = person parameter estimate; RT = runtime (in seconds).

Table 1 includes the *RMSD*s and *RT*s for the main effects and Table 2 includes the eta-squared values for the main and interaction effects.

3.1.1 Accuracy of Item Discrimination Parameter Estimates

The average *RMSD* of item discrimination estimates across all conditions was .256. The number of response categories in the model had a large effect on the accuracy of the

estimates, η^2 = .409, with more response categories resulting in more accurate estimation of item discriminations (*RMSD* =.375, .220, & .173 for 2, 4, & 6 response categories, respectively). Post-hoc Tukey tests for this effect indicated that all pairs of *RMSD* mean differences were statistically significant at α = 0.01. A mean plot for the main effect of number of responses categories on item discrimination estimation accuracy is shown in Figure 2.

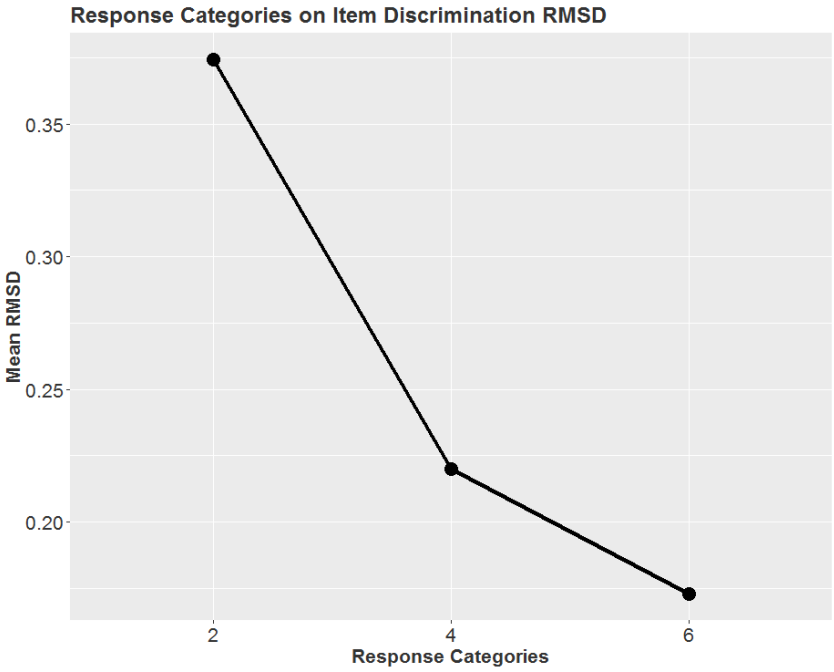

Figure 2 - Mean plot of the main effect of number of response categories on item discrimination estimation accuracy.

3.1.2 Accuracy of Item Location Parameter Estimates

The average *RMSD* of item location estimates across all conditions was .347. The item structure of the model had a large effect on the accuracy of the estimates, η^2 = .323. Namely, the recovery of item location parameters was substantially worse for models with

Table 2 - Eta-squared values for analysis of variance effects

	RMSD				
Effect	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\tau}$	$\hat{\theta}$	RT
Sample size	*.029*	*.017*	*.013*	*<.001*	***.133***
Test length	*.094*	*.093*	*.001*	***.407***	***.272***
Response categories	***.409***	*.027*	*.068*	***.216***	***.348***
Item structure	*.044*	***.323***	***.703***	***.143***	*<.001*
Dimensionality	*.039*	***.102***	*.021*	***.156***	*.063*
Sample size *x* test length	<.001	<.001	<.001	<.001	*.018*
Sample size *x* response categories	*.003*	<.001	*.001*	<.001	*.021*
Test length *x* response categories	*.081*	*.048*	*.007*	*.006*	*.047*
Sample size *x* item structure	<.001	*.002*	*.001*	<.001	*<.001*
Test length *x* item structure	*.019*	*.024*	*.009*	*.005*	*<.001*
Response categories *x* item structure	*.015*	*.011*	*.046*	*.005*	*<.001*
Sample size *x* dimensionality	<.001	<.001	<.001	<.001	*.011*
Test length *x* dimensionality	*.021*	*.040*	*.002*	*.006*	*.007*
Response categories *x* dimensionality	*.022*	*.022*	*<.001*	*.004*	*.029*
Item structure *x* dimensionality	*.002*	*.061*	*.038*	*.018*	*.002*

Note. Values in italics were statistically significant effects at the $p < 0.01$ level; values in bold were effects with η^2 greater than or equal to .10

Table 2 (Continued) - Eta-squared values for analysis of variance effects

	RMSD				
Effect	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\tau}$	$\hat{\theta}$	RT
Sample size *x* test length	<.001	*.002*	*.001*	<.001	*.002*
x response categories					
Sample size *x* test length	<.001	<.001	<.001	<.001	<.001
x item structure					
Sample size *x* response categories	<.001	<.001	<.001	<.001	*<.001*
x item structure					
Test length *x* response categories	*.013*	*.002*	*.002*	*<.001*	*<.001*
x item structure					
Sample size *x* test length	<.001	<.001	<.001	<.001	*.001*
x dimensionality					
Sample size *x* response categories	<.001	<.001	<.001	<.001	*.016*
x dimensionality					
Test length *x* response categories	*.004*	*.004*	<.001	*<.001*	*.004*
x dimensionality					
Sample size *x* item structure	<.001	<.001	<.001	<.001	*<.001*
x dimensionality					
Test length *x* item structure	*.002*	*.006*	<.001	<.001	*<.001*
x dimensionality					

Note. Values in italics were statistically significant effects at the $p < 0.01$ level; values in bold were effects with η^2 greater than or equal to .10

complex item structures (*RMSD* = .457) than for models with simple item structures (*RMSD* = .237).

Table 2 (Continued) - Eta-squared values for analysis of variance effects

	RMSD				
Effect	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\tau}$	$\hat{\theta}$	RT
Response categories *x* item structure	*.003*	*<.001*	*.001*	*<.001*	*<.001*
x dimensionality					
Sample size *x* test length	*.002*	<.001	<.001	<.001	*<.001*
x response categories					
x item structure					
Sample size *x* test length	<.001	<.001	<.001	<.001	*.002*
x response categories					
x dimensionality					
Sample size *x* test length	<.001	<.001	<.001	<.001	<.001
x item structure					
x dimensionality					
Sample size *x* response categories	<.001	<.001	<.001	<.001	*.001*
x item structure					
x dimensionality					
Test length *x* response categories	<.001	*.002*	*<.001*	*<.001*	*<.001*
x item structure					
x dimensionality					
Sample size *x* test length	<.001	<.001	<.001	<.001	*<.001*
x response categories					
x item structure					
x dimensionality					
Model	**.807**	**.790**	**.918**	**.970**	**.980**
(Sum of squares)	(58.68)	(121.78)	(91.17)	(71.132)	(23,684*)

Note. Values in italics were statistically significant effects at the $p < 0.01$ level; values in bold were effects with η^2 greater than or equal to .10

The number of dimensions in the model had a moderate effect on the accuracy of the item location estimates, η^2 = .102. Namely, item locations were more accurately estimated in two-dimensional models (*RMSD* = .285) than in three-dimensional models (*RMSD* = .409).

3.1.3 Accuracy of Subjective Response Category Threshold Estimates

The average *RMSD* of subjective response category threshold estimates across all conditions was .313. The item structure of the model had a large effect on the accuracy of the estimates, η^2 = .703. Namely, the recovery of subjective response category threshold parameters was substantially better for models with complex item structures (*RMSD* = .172) than for models with simple item structures (*RMSD* = .453). This pattern is opposite to that observed for the recovery of item location and person parameters.

3.1.4 Accuracy of Person Parameter Estimates

The average *RMSD* of person parameter estimates across all conditions was .429. Test length had a large effect on the accuracy of the estimates, η^2 = .407. Namely, the recovery of person parameters improved as the test length increased (*RMSD* = .556, .402, & .329 for 10, 20, & 30 items, respectively). Post-hoc Tukey tests indicated that all pairs of *RMSD* mean differences were statistically significant at α = 0.01 for this effect. A mean plot for the main effect of test length on person parameter estimation accuracy is shown in Figure 3.

The number of response categories had a moderate effect on the accuracy of the person parameter estimates, η^2 = .216. Namely, the recovery of person parameters

improved as the number of response categories increased (*RMSD* = .526, .389, & .372 for 2, 4, & 6 response categories, respectively). Post-hoc Tukey tests for this effect indicated that all pairs of *RMSD* mean differences were statistically significant at $\alpha = 0.01$. A mean plot for the main effect of number of response categories on person parameter estimation accuracy is shown in Figure 4.

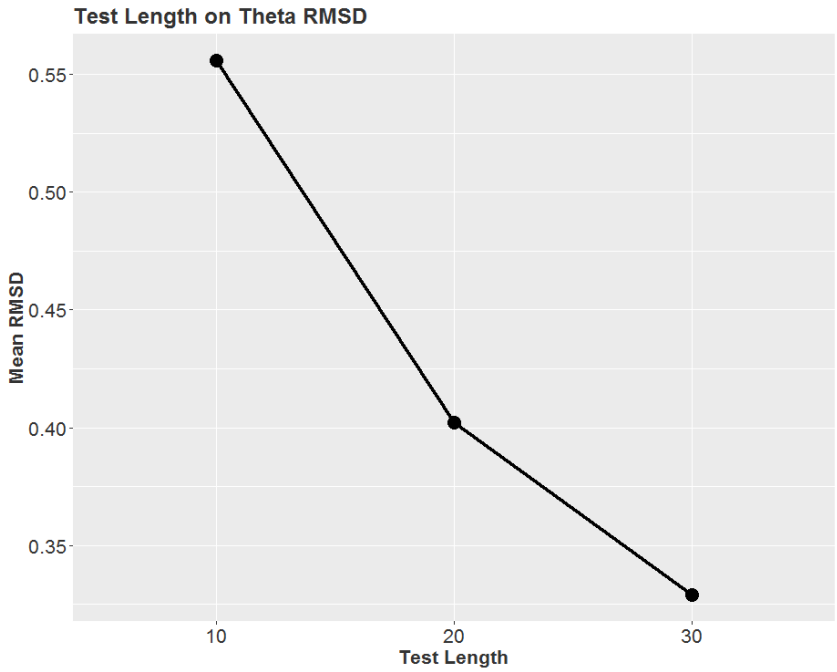

Figure 3 - Mean plot of the main effect of test length on person parameter estimation accuracy.

The item structure of the model had a moderate effect on the accuracy of the person parameter estimates, $\eta^2 = .143$. Namely, the recovery of person parameters was better for models with simple item structures (*RMSD* = .373) than for models with complex item structures (*RMSD* = .485).

The number of dimensions in the model had a moderate effect on the accuracy of the person parameter estimates, $\eta^2 = .156$. Namely, the recovery of person parameters was

better for two-dimensional models (*RMSD* = .370) than for three-dimensional models (*RMSD* = .488).

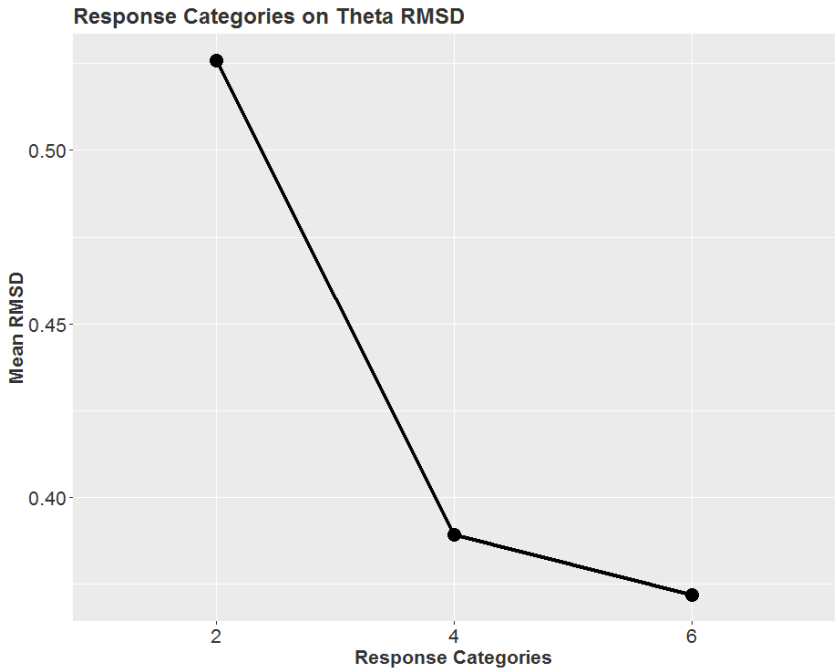

Figure 4 - Mean plot of the main effect of number of response categories on person parameter estimation accuracy.

3.1.5 Runtime of the Modified MH-RM Method

The average length of time for the estimation method to complete (i.e., for convergence criteria to be met) across all conditions was 4,081 seconds (68 minutes). The number of response categories had a large effect on the runtime of the estimation procedure, η^2 = .348. Namely, the runtime increased as the number of response categories increased (*RT* = 1852, 3594, & 5752 for 2, 4, and 6 response categories, respectively). Post-hoc Tukey tests indicated that all pairs of *RT* mean differences for this effect were statistically significant at α = 0.01. A mean plot for the main effect of number of response categories on runtime is shown in Figure 5.

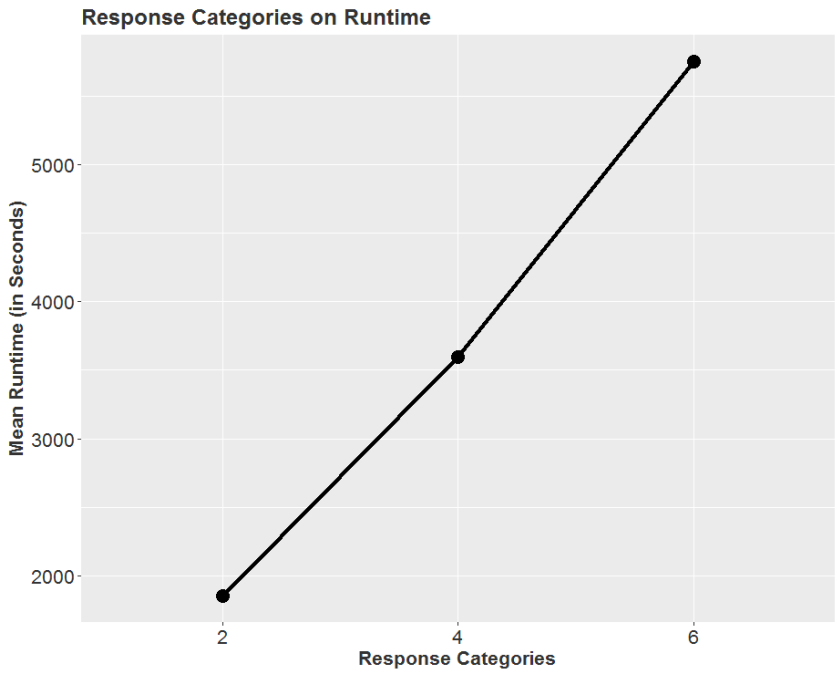

Figure 5 - Mean plot of the main effect of number of response categories on the length of time for the estimation method to complete.

The test length had a large effect on the runtime of the estimation procedure, η^2 = .272. Namely, the runtime increased as the test length increased (*RT* = 1975, 3799, and 5424 for 10, 20, and 30 items respectively). Post-hoc Tukey tests for this effect indicated that all pairs of *RT* mean differences were statistically significant at α = 0.01. A mean plot for the main effect of test length on runtime is shown in Figure 6.

The sample size had a moderate effect on the runtime of the estimation procedure, η^2 = .133. Namely, the runtime increased as the sample size increased (*RT* = 2534, 3717, and 4947 for 1000, 15000, and 2000 persons respectively). Post-hoc Tukey tests indicated that all pairs of *RT* mean differences were statistically significant at α = 0.01 for this effect. A mean plot for the main effect of sample size on runtime is shown in Figure 7.

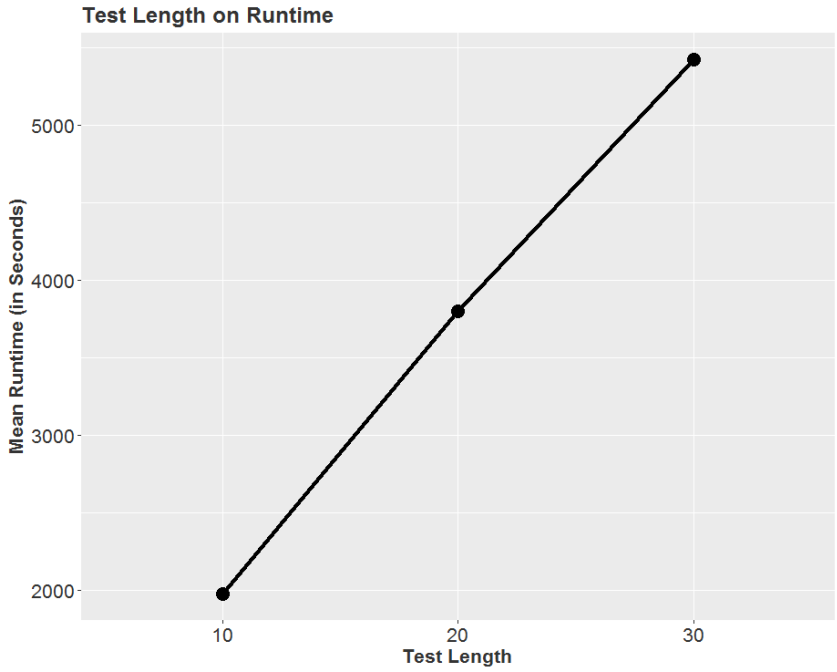

Figure 6 - Mean plot of the main effect of test length on the length of time for the estimation method to complete.

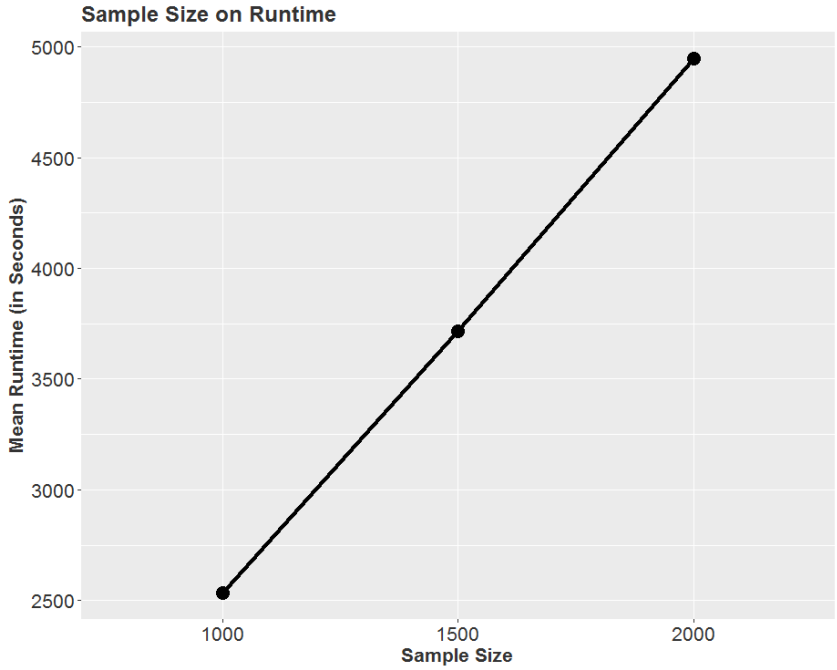

Figure 7 - Mean plot of the main effect of sample size on the length of time for the estimation method to complete.

3.1.6 Item and Person Parameter Standard Errors

Review of the stochastic item parameter standard errors revealed that some of the values were unrealistically large (e.g., >100). Therefore, prior to computing the average standard errors, a check was performed on each standard error value to ensure that the value was contained in the interval [0,10]. If the value was not contained in this interval, it was treated as missing data. Missing data percentages were recorded for each condition. Missing data percentages were not recorded for person parameter standard errors because these values were computed differently (i.e., using the EAP method) from the item parameter standard errors and did not appear to contain any unrealistic values. Average standard errors and missing data percentages for all main effects in the design are reported in Appendix A.

Missing standard errors appeared to occur at the item level. Namely, the standard errors across all parameter types were either missing or present for a given item. Standard errors generally decreased in the expected direction, with increasingly smaller item parameter standard errors observed as the sample size increased, and increasingly smaller person parameter standard errors observed as the test length increased. The one exception was for item location standard errors. As sample size increased, the standard errors did not consistently decrease, indicating that there may not have been a relationship between sample size and item location standard errors. This pattern may have been observed because the smallest sample size in the design was $J = 1000$, and the effect of sample size on item location standard errors may only exist for smaller sample sizes. Standard errors also generally followed the pattern observed in the *RMSD* analysis for item structure and dimensionality. For complex item structure, standard errors were smaller in comparison to

simple item structure for item discriminations and subjective response category thresholds, and larger for item locations and person parameters. Standard errors were smaller for two-dimensional MGGUMs than for three-dimensional MGGUMs.

The missing standard errors seemed to largely be related to item structure and the number of dimensions in the model. Namely, models with simple item structure had approximately 33% missing data for the item parameter standard errors in comparison to approximately 5% of the values for complex item structure models. Further, models with two-dimensions had approximately 10% missing data, whereas models with three-dimensions had approximately 28% missing data.

The missing data also appeared to be related to the number of response categories in the model and the test length. Namely, more response categories was related to less missing data, and longer test length was also related to less missing data.

3.2 Comparison of Modified MH-RM and MMAP Estimation Methods

As described in the method section, a secondary analysis was conducted to directly compare the runtime and estimation accuracy of the modified MH-RM and MMAP methods. Five separate split-plot factorial ANOVAs were conducted to examine the effects of estimation method, test length, sample size, number of dimensions, and item structure on runtime and estimation accuracy. (The number of response categories was held constant at four.) Namely, one univariate split-plot ANOVA was conducted for each of the four parameter types in the MGGUM ($\alpha, \beta, \tau, \& \theta$) and for the runtime measure. Estimation method was treated as a within-subjects factor because both methods were applied to the same response data for each replication in the subset design.

The secondary analysis used the same criteria for interpretation as the primary analysis (i.e., $p < 0.01$ & $\eta^2 \geq 0.10$). However, only the *within-subjects* effects meeting the interpretation criteria were examined in detail because the focus of the secondary analysis was on differences between the modified MH-RM and MMAP methods. For calculating eta-squared values for the within-subjects effects, the total sum of squares value in the denominator was calculated as the total sum of squares of the within-subjects effects including the within-subjects error term . This eta-squared value was denoted as η^2_{wf} to indicate that the value was referring to a within-family calculation of η^2. The results from the analysis are shown in Table 3, which includes the eta-squared values for the main and interaction effects.

3.2.1 Differences in Estimation Accuracy

Model parameters were estimated more accurately with the MMAP method than with the modified MH-RM method. The largest difference in estimation accuracy was observed for item locations (*RMSD* = .337 and .304 for MH-RM and MMAP, respectively), with estimation method accounting for 11% of the variation in item location *RMSD*s.

The second largest difference in estimation accuracy was observed for item discriminations (*RMSD* = .238 & .214 for MH-RM & MMAP, respectively), with estimation method accounting for 25% of the variation in item discrimination *RMSD*s.

Table 3 - Eta-squared within-family values for analysis of variance effects in reduced design

	RMSD				
Effect	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\tau}$	$\hat{\theta}$	RT
Within-Subjects					
Estimation method	***.245***	***.110***	*.081*	.053	***.232***
Sample size	<.001	<.001	.007	.010	.007
x estimation method					
Test length x estimation method	.039	.044	.009	.040	.004
Item structure	.022	.054	.001	.053	*.039*
x estimation method					
Dimensionality	*.069*	.028	.053	.044	***.488***
x estimation method					
Sample size x test length	<.001	<.001	<.001	.007	.007
x estimation method					
Sample size x item structure	.002	<.001	.005	.011	.003
x estimation method					
Test length x item structure	.046	.060	.035	.036	.014
x estimation method					
Sample size x dimensionality	<.001	.013	<.001	.005	.003
x estimation method					
Test length x dimensionality	.021	.019	.042	.040	*.022*
x estimation method					
Item structure x dimensionality	.030	.003	.003	.048	*.027*
x estimation method					

Note. Values in italics were statistically significant effects at the $p < 0.01$ level; values in bold were effects with η^2_{wf} greater than or equal to .10

Table 3 (Continued) - Eta-squared within-family values for analysis of variance effects in reduced design

	RMSD				
Effect	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\tau}$	$\hat{\theta}$	RT
Within-Subjects					
Sample size x test length x item structure x estimation method	.003	<.001	.001	.007	<.001
Sample size x test length x dimensionality x estimation method	<.001	.010	.016	.004	<.001
Sample size x item structure x dimensionality x estimation method	<.001	.003	.003	.005	<.001
Test length x item structure x dimensionality x estimation method	.011	.013	.003	.034	.013
Sample size x test length x item structure x dimensionality x estimation method	<.001	.003	.006	.004	<.001
Replications (sample size test length item structure dimensionality) x estimation method	.513	.639	.731	.597	.138
Model	**.487**	**.361**	**.269**	**.403**	**.862**
(Sum of squares within family)	(.095)	(.412)	(.027)	(.026)	2,385*

Note. Values in italics were statistically significant effects at the $p < 0.01$ level; values in bold were effects with η^2_{wf} greater than or equal to .10

3.2.2 Differences in Runtime

Estimation method had a moderate effect on runtime, η^2_{wf} = .232. The modified MH-RM method completed in less time (RT = 3721) than the MMAP method (RT = 7441). Estimation method also affected runtime through an interaction with the number of

dimensions in the model. Namely, the interaction between estimation method and number of dimensions had a large effect on the runtime of the estimation procedure, η^2_{wf} = .488. For two-dimensional models, MMAP estimation completed in less time (RT = 1641) than MH-RM estimation (RT = 3319). Although for three-dimensional models, the opposite pattern was observed. Namely, MH-RM estimation completed in substantially less time (RT = 4124) than MMAP estimation (RT = 13241). A mean plot for the interaction effect of number of dimensions by estimation method on runtime is shown in Figure 8.

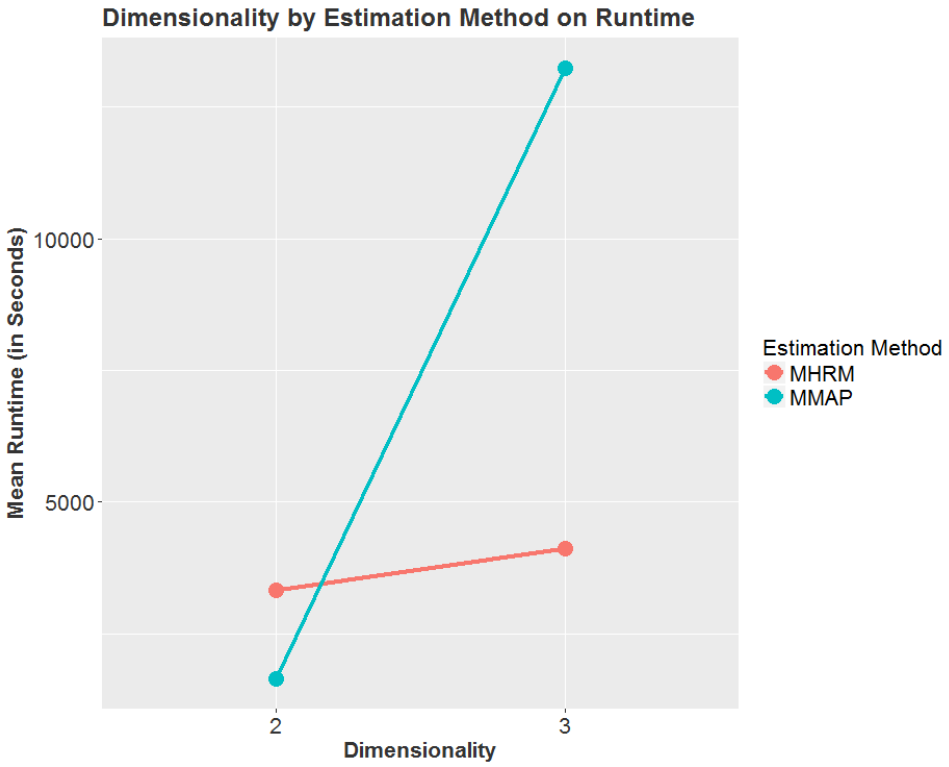

Figure 8 - Mean plot of the interaction effect of number of dimensions by estimation method on runtime.

A simple main effects analysis was conducted to further elucidate which runtime mean differences were driving the interaction effect. Conditioning on estimation method, two independent-samples *t*-tests were conducted at α = .01/2 = .005, to examine

differences in runtime between two- and three-dimensional models. For the MMAP method, the mean difference in runtime between the two- and three-dimensional models was statistically significant. However, the mean difference was **not** statistically significant between the two- and three-dimensional models for the MH-RM method (p = .134). Conditioning on number of dimensions, two paired-samples t-tests were conducted at α = .01/2 = .005, to examine differences in runtime between the MH-RM and MMAP methods. Both tests were statistically significant, indicating that for two-dimensional models, the mean runtime for the MMAP method was statistically less than the mean runtime for the MH-RM method, and for three-dimensional models, the mean runtime for the MMAP method was statistically greater than the mean runtime for the MH-RM method.

3.2.3 Item and Person Parameter Standard Errors

The factors in the reduced design appeared to have a similar effect on the standard errors as the factors in the full design had on the standard errors. Additionally, the estimation method appeared to have an effect on the standard errors, with standard errors from the MMAP method somewhat smaller than standard errors from the modified MH-RM method. Further, the MMAP method often had fewer missing standard errors than the modified MH-RM method, although the percentages of missing data were comparable between the two methods across all cells in the reduced design.

3.3 Empirical Data Application

A real data analysis examined the utility of the modified MH-RM algorithm for estimating MGGUM parameters in practical situations. The analysis was performed on 1237 graded disagree-agree responses to 24 emotion stimuli. Emotions represented by the

stimuli included *afraid, angry, calm, disgusted, happy, neutral, sad,* & *surprised*. There were three stimuli for each of the eight emotions.

Previous MGGUM and multidimensional scaling solutions (Roberts & Sparks, 2015; Sparks, 2015) suggested these data are three-dimensional. The three hypothesized dimensions were *valence*, *activation*, and *potency*. This interpretation was based on the circumplex theory of emotion developed by Schlosberg (1941; 1952). The original theory mapped emotional states on the circumference of a circle, with coordinates determined by two orthogonal dimensions: pleasure and activation. More recently, the two primary dimensions in the circumplex model have been described as *valence* and *activation* (Russell, 1980). Valence is the extent to which an emotion is pleasant (e.g., *happy*) or unpleasant (e.g., *sad*). Activation is the extent to which an emotion is energizing (e.g., *surprised*) or de-energizing (e.g., *calm*). However, some emotions are not well distinguished by these two dimensions. For example, fear and anger both have negative valence and high activation. Therefore, a third dimension, *potency*, has been proposed (Russell & Mehrabian, 1977) to distinguish between emotions that are not well separated in the circumplex model. Potency is the extent to which a person feels empowered (e.g., *angry*) or disempowered (e.g., *afraid*). These three dimensions correspond with the three primary dimensions Osgood, Succi, and Tannenbaum (1957) derived from a factor analysis on stimulus ratings.

In the current study, both two and three-dimensional MGGUMs were fit to the emotion data using the modified MH-RM method. The solutions were assessed with the Bayesian information criterion (BIC; Schwarz, 1978), and a pseudo-Bayes factor (PsBF) based on the conditional predictive ordinate (CPO; Gelfand, Dey, & Chang, 1992).

3.3.1 Relative Model Fit Indices

The BIC is a function of relative model fit, measured by the term $-2\ln(L)$, and model complexity, measured by the term $2p\ln(N)$:

$$\text{BIC} = -2\ln(L) + 2p\ln(N), \tag{21}$$

where *L* denotes the likelihood, *p* denotes the number of estimated parameters in the model, and *N* denotes the sample size. The model with the lowest BIC value is preferred.

PsBFs are often used for assessing relative model fit in Bayesian applications. Geisser and Eddy (1979) described a PsBF as the ratio between the *estimated* marginal likelihoods for Model A and Model B:

$$\text{PsBF} = \frac{\hat{L}(\text{data}|\text{Model A})}{\hat{L}(\text{data}|\text{Model B})} \tag{22}$$

A PsBF > 1 suggests Model A is preferred, whereas a PsBF < 1 suggests Model B is preferred. The CPO can be used as an estimate of the marginal likelihood, $L(\text{data}|\text{model})$. For item response, z_{ij}, the CPO is defined as:

$$\text{CPO}_{ij}^{-1} = \frac{1}{T}\sum_{t=1}^{T} 1/\, f(z_{ij} \mid \Psi_t, \Theta_t), \tag{23}$$

where *T* is the number of samples taken to compute the index after the modified MH-RM procedure has converged, and $f(z_{ij} \mid \Psi_t, \Theta_t)$ is the likelihood of the observed response given the model parameters on the *t*th iteration of the MH-RM algorithm. In this study, *T* = 30 samples were taken for calculating the CPO index. The two-dimensional model was

specified as Model A, and the three-dimensional model was specified as Model B. The PsBF was then calculated according to the following formula:

$$\text{PsBF} = \frac{\prod_{j=1}^{J}\prod_{i=1}^{I}\text{CPO}_{ij}(\text{Model A})}{\prod_{j=1}^{J}\prod_{i=1}^{I}\text{CPO}_{ij}(\text{Model B})}. \quad (24)$$

3.3.2 Analysis

Both relative model fit indices favored the three-dimensional solution (BIC = 61,549.17, PsBF < .001) over the two-dimensional solution (BIC = 63,485.83). Therefore, the three-dimensional solution was selected for interpretation. The item discrimination and item location parameter estimates for this solution are given in Table 4. Multidimensional discrimination values (MDISC; Reckase & McKinley, 1991) are also included in the table. MDISC provides an index of each item's overall discriminating power. MDISC values ranged from 1.148 (*Angry3*) to 4.815 (*Happy1*).

Dimensions were interpreted by identifying the most discriminating items on each dimension and then evaluating the corresponding item location estimates. On the first dimension, the most discriminating items were in the *Happy* and *Neutral* stimulus sets. For the *Happy set*, $\hat{\alpha}_{Happy1,1}$ = 3.334 and $\hat{\hat{\delta}}_{Happy1,1}$ = 2.902. For the *Neutral set*, $\hat{\alpha}_{Neutral3,1}$ = 2.117 and $\hat{\delta}_{Neutral3,1}$ = .287. This seemed to suggest that the first dimension measured valence. As expected, items in the *Sad* set had negative item location estimates (e.g., $\hat{\delta}_{Sad3,1}$ = –4.543) on the first dimension. Notably, the *Surprised* items did not load on the first dimension.

This seems reasonable because the feeling of surprise is not intrinsically pleasant or unpleasant (and may be either pleasant or unpleasant depending on the subjective experience of the person).

On the second dimension, the most discriminating items were in the *Calm*, *Neutral*, and *Happy* sets. The item location estimates for the *Calm* and *Neutral* items were strongly negative, whereas the item location estimates for the *Happy* items were strongly positive. This seemed to suggest that the second dimension measured activation. However, according to the circumplex model, the *Happy* items should have location estimates closer to the midpoint of the scale. Further, the *Sad* items had location estimates that were strongly negative. According to the circumplex model, the *Sad* items should have location estimates closer to the midpoint as well. Although, the *Surprised* items had the highest positive item locations, as would be expected if the underlying dimension measured activation. Interestingly, the *Afraid*, *Disgusted*, and *Angry* items did not load on the second dimension. The circumplex model suggests that *Afraid* and *Angry* are both emotions with high activation, and therefore these items should have loaded on the second dimension if this dimension measured activation. The high, positive item location estimates for the *Happy* and *Surprised* items suggest that the dimension might measure positive activation, or a composite of activation and valence.

Figure 9 shows a bivariate plot of dimension 1 (*valence*) by dimension 2 (*activation composite*). The plot appears to show a portion of the emotion circumplex. Namely, the *Sad*, *Neutral*, *Calm*, *Happy*, and *Surprised* item sets appear to be in the approximately correct relative position on the circumference of the circle. The *Afraid* and *Angry* item sets appear to be in the correct relative position on the valence dimension, although should be

located higher on the activation dimension. (Recall that these item sets did not load on the activation dimension). If the solution were rotated clockwise about 45 degrees, it would correspond well with the traditional interpretation of the valence and activation dimensions.

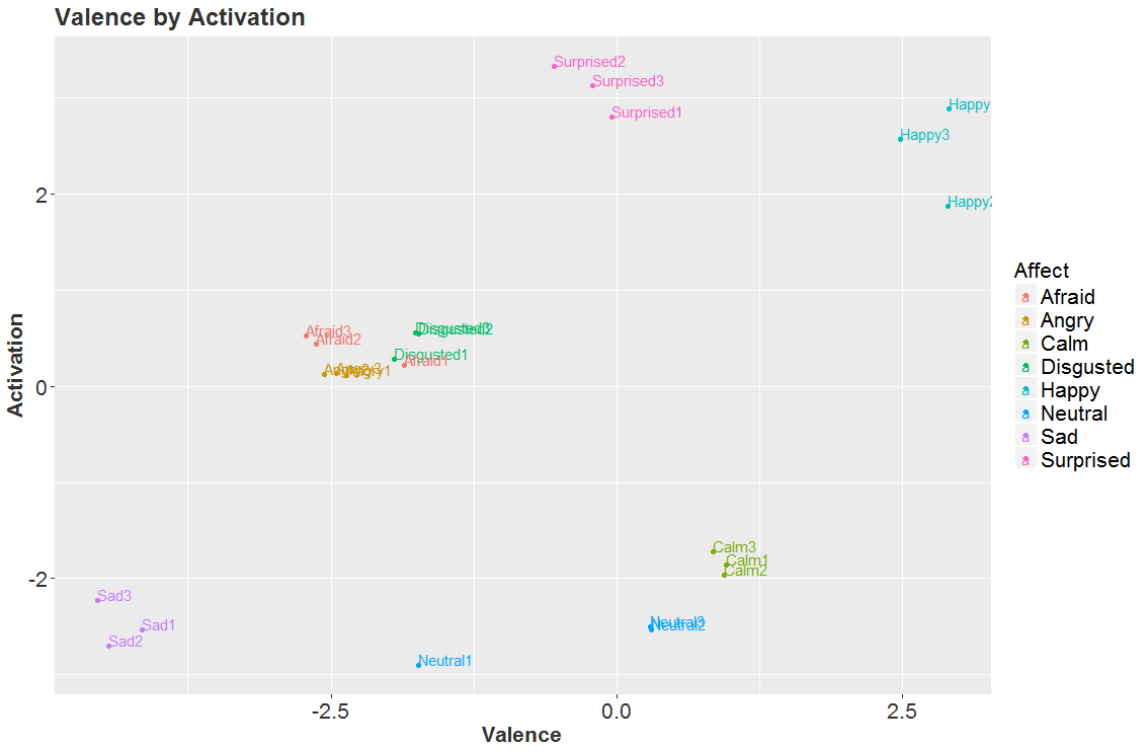

Figure 9 - Bivariate plot of item locations for dimension 1 (valence) and dimension 2 (activation composite).

On the third dimension, the most discriminating items were in the *Afraid*, *Happy*, and *Surprised* sets. The *Afraid* and *Surprised* sets had positive item location estimates, whereas the *Happy* set had negative item location estimates. Further, the *Calm* items also had negative item locations. It is unclear what this dimension measures, although the dimension does appear to distinguish between *Angry* items

$(\hat{\delta}_{Angry1,3} = 1.503,\ \hat{\delta}_{Angry2,3} = 1.585, \& \hat{\delta}_{Angry3,3} = 1.769)$ and *Afraid* items

Table 4 - MGGUM item discrimination and item location parameter estimates for emotion stimuli

Item	*MDISC*	$\hat{\alpha}_{i1}$	$\hat{\alpha}_{i2}$	$\hat{\alpha}_{i3}$	$\hat{\delta}_{i1}$	$\hat{\delta}_{i2}$	$\hat{\delta}_{i3}$
Angry1	*1.292*	0.757	0.000	1.048	-2.359	0.112	1.503
Angry2	*1.293*	0.750	0.000	1.052	-2.554	0.121	1.585
Angry3	*1.148*	0.732	0.000	0.884	-2.450	0.137	1.769
Calm1	*2.346*	1.717	1.486	0.588	0.958	-1.856	-0.916
Calm2	*2.490*	1.708	1.630	0.792	0.940	-1.967	-0.200
Calm3	*2.403*	1.471	1.776	0.674	0.840	-1.723	-0.792
Disgusted1	*1.469*	0.775	0.000	1.248	-1.946	0.284	0.939
Disgusted2	*1.182*	0.595	0.000	1.021	-1.737	0.547	0.818
Disgusted3	*1.520*	0.680	0.000	1.359	-1.758	0.557	0.939
Afraid1	*2.174*	0.812	0.001	2.017	-1.857	0.217	0.584
Afraid2	*2.603*	0.824	0.000	2.469	-2.628	0.436	0.554
Afraid3	*2.575*	0.747	0.003	2.464	-2.716	0.529	0.545
Happy1	*4.815*	3.334	1.720	3.018	2.902	2.882	-0.388
Happy2	*4.040*	2.820	1.712	2.331	2.895	1.873	-0.607
Happy3	*3.442*	2.601	1.289	1.849	2.478	2.573	-0.572
Neutral1	*1.678*	0.831	1.174	0.864	-1.737	-2.899	-0.057
Neutral2	*2.690*	1.627	1.870	1.043	0.298	-2.536	0.044
Neutral3	*3.516*	2.117	2.327	1.568	0.287	-2.498	-0.333

Table 4 (Continued) - MGGUM item discrimination and item location parameter estimates for emotion stimuli

Item	*MDISC*	$\hat{\alpha}_{i1}$	$\hat{\alpha}_{i2}$	$\hat{\alpha}_{i3}$	$\hat{\delta}_{i1}$	$\hat{\delta}_{i2}$	$\hat{\delta}_{i3}$
Sad1	*1.584*	0.874	0.519	1.215	-4.153	-2.528	0.504
Sad2	*1.593*	0.818	0.477	1.281	-4.442	-2.706	0.601
Sad3	*1.737*	0.781	0.614	1.425	-4.543	-2.232	0.328
Surprised1	*2.079*	0.000	0.448	2.030	-0.044	2.798	0.846
Surprised2	*2.757*	0.000	0.323	2.738	-0.553	3.328	0.782
Surprised3	*2.923*	0.001	0.518	2.877	-0.213	3.131	0.838

($\hat{\delta}_{Afraid1,3} = 0.584,\ \hat{\delta}_{Afraid2,3} = 0.554, \& \hat{\delta}_{Afraid3,3} = 0.545$) . These two item sets were not well distinguished on dimensions one ($\hat{\delta}_{Angry1,1} = -2.359,\ \hat{\delta}_{Angry2,1} = -2.555,$ $\& \hat{\delta}_{Angry3,1} = -2.450$; $\hat{\delta}_{Afraid1,1} = -1.857,\ \hat{\delta}_{Afraid2,1} = -2.628, \& \hat{\delta}_{Afraid3,1} = -2.716$) or two ($\hat{\delta}_{Angry1,2} = 0.112,\ \hat{\delta}_{Angry2,2} = 0.121, \& \hat{\delta}_{Angry3,2} = 0.138$; $\hat{\delta}_{Afraid1,2} = 0.217,\ \hat{\delta}_{Afraid2,2} = 0.437,\ \ \& \hat{\delta}_{Afraid3,2} = 0.529$) . This provides evidence that the dimension measures potency, because potency distinguishes some negative emotions from each other (Russell & Mehrabian, 1977). Dimension three was interpreted as a composite that measured potency and possibly valence and activation to some extent as well. Figure 10 shows dimension 1 (valence) by dimension 3 (potency composite), and Figure 11 shows dimension 2 (activation composite) by dimension 3 (potency composite).

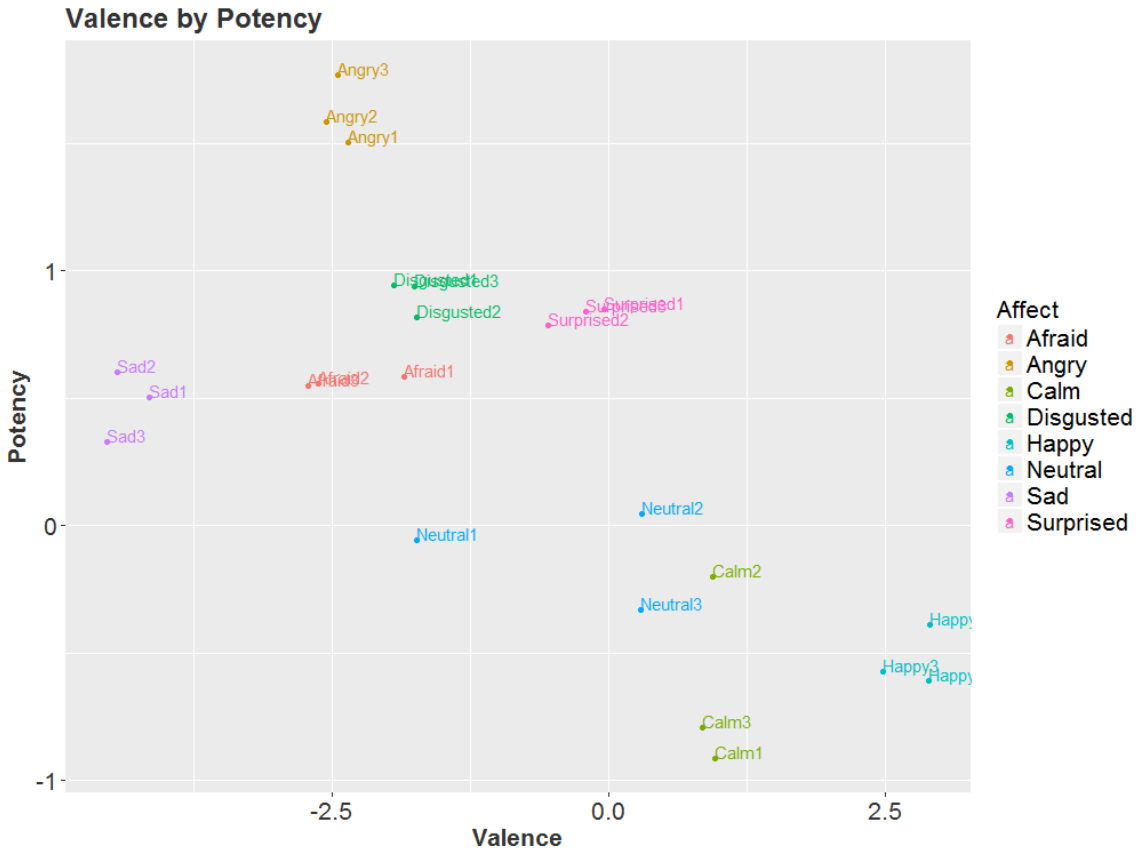

Figure 10 - Bivariate plot of item locations for dimension 1 (valence) and dimension 3 (potency composite).

The bivariate plot of activation and potency (Figure 11) shows that the *Angry* and *Afraid* items are not distinguished on the activation dimension, although they are distingushed on the potency dimension. Finally, in terms of face validity, the three stimuli for each emotion appeared to cluster together in each of the bivariate plots. If the similar items had not clustered together, it would have raised questions about the quality of the stimuli or the ability of the model to accurately estimate item locations.

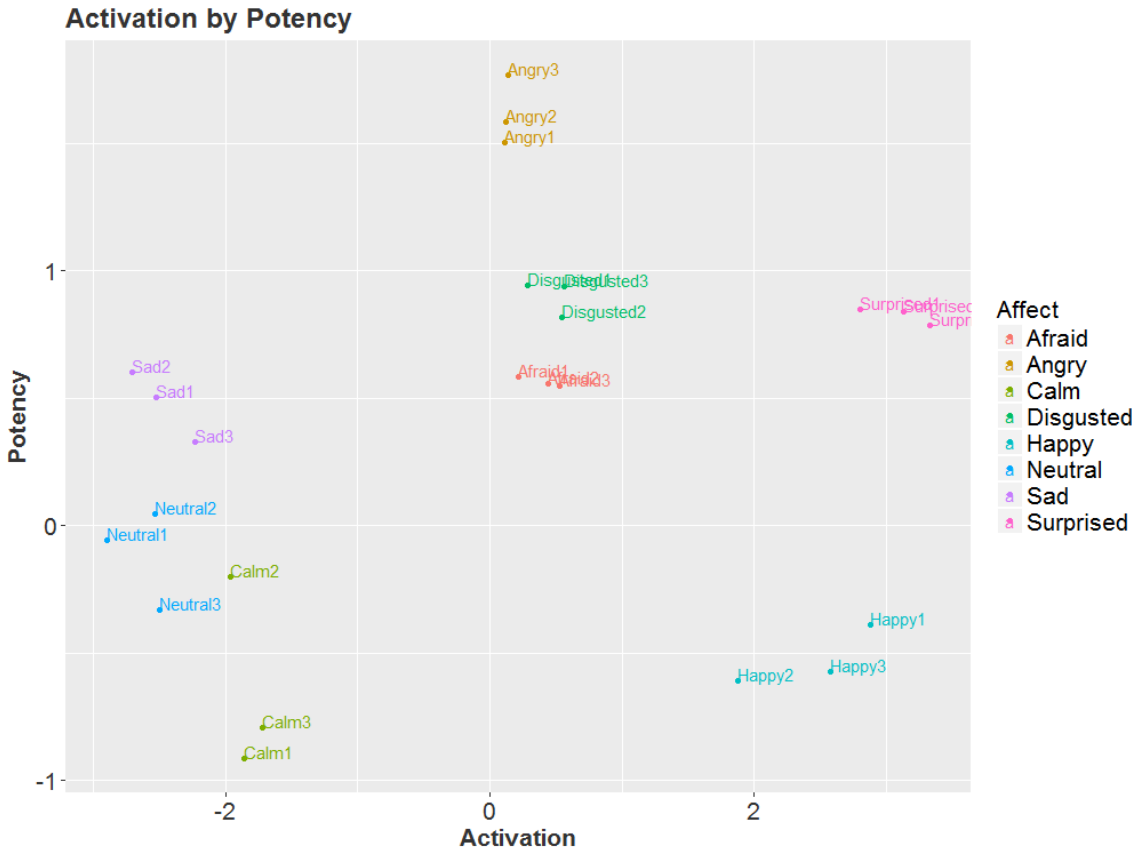

Figure 11 - Bivariate plot of item locations for dimension 2 (activation composite) and dimension 3 (potency composite).

CHAPTER 4. DISCUSSION

The primary goal of the study was to examine the speed and estimation accuracy of the modified MH-RM. Results from the parameter recovery study indicated that the modified MH-RM was faster than the MMAP method for three dimensions, and that differences in estimation accuracy between the modified MH-RM and MMAP methods were small.

The runtime patterns observed for two- and three-dimensional models indicated that increasing the number of dimensions in the model led to small, linear increases in runtime for the modified MH-RM, and relatively large, exponential increases in runtime for the MMAP method. For the modified MH-RM method, a three-dimensional model was estimated in approximately 68.73 minutes, a 24% increase in runtime over estimation of a two-dimensional model (55.32 minutes). For the MMAP method, a three-dimensional model was estimated in approximately 220.68 minutes, an 807% increase in runtime over estimation of a two-dimensional model (27.35 minutes). This pattern suggests that as the number of dimensions increases, the MMAP method will eventually become intractable given current computer memory and processing speed limitations. Further, if a solution could be obtained for high-dimensional models, the MMAP method may be impractical for applied researchers wanting a solution in hours or days, rather than weeks. The modified MH-RM is an attractive option for the estimation of high-dimensional models because increasing the number of dimensions in the model leads to minimal increases in runtime.

The *RMSD* patterns indicated that item discrimination and location parameters were estimated more accurately with the MMAP method by a small margin over the modified MH-RM method, although subjective response category threshold and person parameters were estimated at essentially the same level of accuracy for both methods. For person parameter estimation, this is not a surprising result because the EAP method was used to estimate person parameters in both the modified MH-RM and MMAP conditions. However, the EAP method did condition on the item parameter estimates obtained from the MH-RM and MMAP solutions (i.e., item parameters were used for computing the likelihood at each quadrature point). This indicated that differences in the accuracy of the item parameter estimates between the two estimation methods had a negligible effect on the accuracy of person parameter estimates.

As the focus of psychological research is often on the measurement of individual differences, the estimation accuracy of person parameters is arguably of higher importance than the estimation accuracy of item parameters. Further, item parameters are typically measured more accurately than person parameters in item response theory models because item responses from J persons are used to estimate item parameters, whereas item responses from I items are used to estimate person parameters (with J often greatly exceeding I).

4.1 The Speed-Accuracy Tradeoff

Both the modified MH-RM and MMAP methods can be specified to prioritize either speed or accuracy. For the modified MH-RM method, estimation accuracy can be improved (at the cost of speed) by increasing the number of burn-in cycles, decreasing the

rate at which the Robbins-Monro gain constant decreases, and/or increasing the number of theta samples that are stochastically imputed at each stage. Similarly, for the MMAP method, estimation accuracy can be improved (at the cost of speed) by increasing the number of quadrature points for each dimension and decreasing the maximum difference in parameter estimates allowed between adjacent cycles (i.e., the tolerance) for the convergence criteria to be met. If the MMAP method were used for estimating higher-dimensional models (e.g., 5 dimensions), the number of quadrature points per dimension would need to be reduced for a solution to be obtained in a timely manner. However, decreasing the number of quadrature points would likely also decrease the estimation accuracy. Given the limitations of the MMAP procedure for estimating higher-dimensional models, the MH-RM method seems preferable for estimating models with four or more dimensions. For three-dimensional models, the method selected might depend on the goals of the particular study. Finally, for two-dimensional models, the MMAP procedure is clearly preferred on account of both faster runtime and better estimation accuracy.

4.2 Data Demands for Improving Estimation Accuracy

Increasing the number of response categories in the model led to improved estimation of item discrimination parameters and person parameters, with larger improvements observed for the increase from 2 to 4 response categories than for the increase from 4 to 6 response categories. However, increasing the number of response categories in the model also led to longer runtimes. Although in practice, longer runtimes may be a small price to pay for the improvements in estimation accuracy of item discrimination and person parameters. The patterns observed in this study for the effect of

number of response categories on item discrimination and person parameter estimation accuracy were similar to those reported in Thompson (2014).

As expected, increasing the test length led to improved estimation of person parameters, with larger improvements observed for the increase from 10 to 20 items than for the increase from 20 to 30 items. These results corroborated the results presented in the Thompson (2014) study.

Finally, in the current study, complex item structure led to improved estimation of subjective response category threshold parameters, when compared to simple item structure. However, complex item structure led to decreased estimation accuracy of item location and person parameters. These findings also corroborated the findings reported in Thompson (2014).

4.3 Scientific Importance of the Current Study

Initial attempts to estimate the MGGUM with the MH-RM algorithm resulted in severe misestimation of some item parameters, although replacement of the Newton-Raphson step for updating the item parameter estimates with the L-BFGS-B method markedly improved estimation accuracy. This modification to the MH-RM algorithm appears to have stabilized the method in two ways. First, the L-BFGS-B method did not require inversion of a Hessian matrix, a computation that was problematic because the inverse of the Hessian did not always exist (or could not be found). Second, the method updated item parameter estimates multiple times on each stage of the algorithm (until the specified convergence criteria was met), in contrast to the Newton-Raphson single update of the item parameters.

In previous MGGUM research, the inversion of the Hessian matrix has been problematic. Thompson (2014) encountered this problem and settled on an approach that inverted sections of the Hessian for different sets of item parameters, rather than inverting the full Hessian with respect to all item parameters. In the current study, the inverse of the full Hessian was approximated using the L-BFGS-B method for both the MMAP and modified MH-RM methods.

Further, the current study contributed to the development of the MGGUM in the following technical ways: (a) new regression equations were found for obtaining threshold start values and hyperparameters for threshold prior distributions; (b) rescaling constants were found for rescaling threshold true parameters in complex structure conditions to ensure that all response categories were used in the simulation study; (c) first partial derivatives with respect to item parameters were derived and developed into C++ functions for improving the speed and accuracy of the maximization step of the modified MH-RM and MMAP procedures;[13] and (d) second partial derivatives with respect to item parameters were derived and developed into C++ functions for computing all elements of the Hessian matrix, which was then used in the procedure for obtaining item parameter standard errors.[14]

Finally, the modified MH-RM and MMAP methods used in the current study were implemented within R package *mirt*, and will be released to the general public. This will

[13] Note that Thompson (2014) had previously derived the first partial derivatives for the MGGUM, although she used a different formulation and implemented the derivatives in FORTRAN
[14] Previous solutions approximated the Hessian from the product of the derivatives.

give applied researchers a convenient means for conducting multidimensional analyses on preference, satisfaction, or self-similarity data.

4.4 Alternative Estimation Methods

The current study was motivated by the inefficiency of rectangular quadrature estimation in the expectation step of the traditional EM algorithm as the number of dimensions in the model increases. The modified MH-RM method overcame the efficiency limitations of rectangular quadrature by using stochastic imputation to update person parameter estimates in the expectation step. The modified MH-RM method therefore provides a viable alternative to EM for higher-dimensional models because the stochastic imputation approach increases runtime as a linear function of the number of dimensions in the model, as opposed to the exponential function observed with rectangular quadrature. However, other alternative estimation methods might also be explored for overcoming the inefficiency of rectangular quadrature for higher-dimensional models.

Three potential methods are adaptive quadrature (Schilling & Bock, 2005), Monte Carlo quadrature with Gibbs sampling (Meng & Schilling), or MCMC (Patz & Junker, 1999). The first two methods use more efficient techniques than rectangular quadrature for updating the person parameter estimates in the expectation step of the EM algorithm, whereas the MCMC method uses a completely different estimation paradigm than EM. Adaptive quadrature is attractive because the method has been shown to work with as few as two quadrature points. Although it is unclear how well the method would hold up for high-dimensional, non-compensatory models.

Monte Carlo quadrature with Gibbs sampling is another method that might be explored, although Meng and Schilling found that the marginal log-likelihood often decreased across EM cycles. This suggests that the method may not find the optimal solution, and also that convergence criteria may be difficult to specify. Finally, MCMC might be a promising estimation method for high-dimensional MGGUMs. Thompson (2014) reported that the MCMC method took days to complete, although she used the general-purpose WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) software. Direct implementation of the likelihood and proposal distributions in an efficient language such as FORTRAN or C++ might substantially improve the efficiency of the MCMC method.

4.5 Future Directions

Although the evaluation of item parameter standard errors was not a primary focus for the current study, it was observed that some study conditions led to large percentages of missing data (i.e., standard errors with unrealistic values). Namely, the percentage of missing data substantially increased for models with simple item structure. Further, models with fewer items, fewer response categories, and more dimensions also had higher percentages of missing data. This might have resulted from problems related to the inversion of the Hessian matrix in the computation of the standard errors. Future research could investigate the factors contributing to the missing data and could develop a method for robust estimation of item parameter standard errors. One promising approach might be to estimate the inverse Hessian using the same equations used in the L-BFGS-B method. Another potential avenue might be to examine recovery of MGGUM item parameter standard errors using the empirical cross-product or supplemented EM method. Paek and

Cai (2014) found that both of these methods showed good recovery of standard errors for two-dimensional 2PL models.

The modification of the MH-RM method in the current study led to substantial improvements in item parameter estimation over traditional MH-RM. Future research could explore the extent to which these modifications might improve the estimation of other MIRT models that have been estimated with traditional MH-RM. For example, Chalmers and Flora (2014) found that parameters in non-compensatory models were generally not recovered accurately with the MH-RM method. However, the modified MH-RM method developed in the current study might improve estimation of those models.

APPENDIX A. AVERAGE STANDARD ERRORS AND MISSING DATA PERCENTAGES FOR MAIN EFFECTS IN SIMULATION DESIGN

Table 5 - Average standard errors of parameter estimates by condition

	Stochastic						EAP
Effect	$SE_{\hat{\alpha}}$	*missing*	$SE_{\hat{\delta}}$	*missing*	$SE_{\hat{\tau}}$	*missing*	$SE_{\hat{\theta}}$
Sample Size							
1000	.158	*19%*	.214	*18%*	.150	*19%*	.507
1500	.136	*19%*	.213	*18%*	.133	*19%*	.513
2000	.125	*19%*	.224	*18%*	.127	*19%*	.514
Test Length							
10	.135	*24%*	.227	*23%*	.140	*24%*	.663
20	.141	*18%*	.216	*18%*	.135	*18%*	.480
30	.144	*15%*	.208	*14%*	.135	*15%*	.391
Response Categories							
2	.207	*25%*	.230	*24%*	.158	*25%*	.630
4	.127	*17%*	.216	*16%*	.128	*17%*	.474
6	.085	*15%*	.205	*14%*	.124	*15%*	.430

Note. $SE_{\hat{\alpha}}$ = standard error of the item discrimination estimate; $SE_{\hat{\delta}}$ = standard error of the item location estimate; $SE_{\hat{\tau}}$ = standard error of the subjective response category threshold estimate; $SE_{\hat{\theta}}$ = standard error of the person parameter estimate; *Stochastic* = stochastic approximation method; *EAP* = expected a posteriori method; *missing* = percentage of SEs with values less than zero or greater than 10.

Table 5 (Continued) - Average standard errors of parameter estimates by condition

	Stochastic						EAP
Effect	$SE_{\hat{\alpha}}$	*missing*	$SE_{\hat{\delta}}$	*missing*	$SE_{\hat{\tau}}$	*missing*	$SE_{\hat{\theta}}$
Item Structure							
Complex	.105	*5%*	.219	*5%*	.089	*5%*	.566
Simple	.174	*33%*	.215	*32%*	.184	*33%*	.456
Dimensionality							
2	.128	*10%*	.156	*9%*	.133	*10%*	.450
3	.151	*28%*	.278	*27%*	.140	*28%*	.573
Overall	.137	*19%*	.217	*18%*	.137	*19%*	.511

Note. $SE_{\hat{\alpha}}$ = standard error of the item discrimination estimate; $SE_{\hat{\delta}}$ = standard error of the item location estimate; $SE_{\hat{\tau}}$ = standard error of the subjective response category threshold estimate; $SE_{\hat{\theta}}$ = standard error of the person parameter estimate; *Stochastic* = stochastic approximation method; *EAP* = expected a posteriori method; *missing* = percentage of SEs with values less than zero or greater than 10.

APPENDIX B. FIRST AND SECOND PARTIAL DERIVATIVES OF THE COMPLETE DATA LOG LIKELIHOOD WITH RESPECT TO ITEM PARAMETERS

The complete data log likelihood for the MGGUM is:

$$\ln(L) = \sum_{j=1}^{J}\sum_{i=1}^{I} \ln\left(P\left[Z_{ji}\right]\right)$$

$$= \sum_{j=1}^{J}\sum_{i=1}^{I}\left[\begin{array}{l} \sum_{k=0}^{z}\psi_{ik} + \ln\left(\exp\left[z\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right] + \exp\left[(M-z)\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right]\right) \\ -\ln\left(\sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}\psi_{ik}\right)\left(\exp\left[w\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right] + \exp\left[(M-w)\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right]\right)\right)\right) \end{array}\right]$$

It is easier to compute the partial derivatives by including two dummy variables, $U_{z_{jik}}$ and $U_{w_{jik}}$, in the likelihood:

$$\ln(L) = \sum_{j=1}^{J}\sum_{i=1}^{I}\left[\begin{array}{l} \sum_{k=0}^{z}U_{z_{jik}}\psi_{ik} + \ln\left(\exp\left[z\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right] + \exp\left[(M-z)\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right]\right) \\ -\ln\left(\sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp\left[w\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right] + \exp\left[(M-w)\sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}\right]\right)\right)\right) \end{array}\right]$$

where $U_{z_{jik}} = 1$ for $0 < k \le z$, and $U_{z_{jik}} = 0$ otherwise; and

$U_{w_{jik}} = 1$ for $0 < k \le w$, and $U_{w_{jik}} = 0$ otherwise.

Moreover, the symbol Δ is used to denote the distance term in the likelihood:

$$\ln(L) = \sum_{j=1}^{J}\sum_{i=1}^{I}\left[\begin{array}{l} \sum_{k=0}^{z}U_{z_{jik}}\psi_{ik} + \ln\left(\exp\left[z\Delta\right] + \exp\left[(M-z)\Delta\right]\right) \\ -\ln\left(\sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp\left[w\Delta\right] + \exp\left[(M-w)\Delta\right]\right)\right)\right) \end{array}\right]$$

where $\Delta = \sqrt{\sum_{d=1}^{D}\alpha_{id}^2(\theta_{jd}-\delta_{id})^2}$

The first partial derivative of the complete data log likelihood with respect to δ_{id} is:

$$\frac{\partial \ln(L)}{\partial \delta_{id}} = \sum_{j=1}^{J}\sum_{i=1}^{I}\left[\frac{a'}{a} - \frac{b'}{b}\right]$$

Where

$$a = \exp(z\Delta) + \exp((M-z)\Delta)$$

$$a' = \frac{z\alpha_{id}^2(\delta_{id}-\theta_{jd})\left[\exp(z\Delta)\right] + (M-z)\alpha_{id}^2(\delta_{id}-\theta_{jd})\left[\exp((M-z)\Delta)\right]}{\Delta}$$

$$b = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp(w\Delta) + \exp((M-w)\Delta)\right)\right)$$

and

$$b' = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\frac{w\alpha_{id}^2(\delta_{id}-\theta_{jd})\left[\exp(w\Delta)\right] + (M-w)\alpha_{id}^2(\delta_{id}-\theta_{jd})\left[\exp((M-w)\Delta)\right]}{\Delta}\right)\right).$$

The first partial derivative of the complete data log likelihood with respect to α_{id} is:

$$\frac{\partial \ln(L)}{\partial \alpha_{id}} = \sum_{j=1}^{J}\sum_{i=1}^{I}\left[c' + \frac{a'}{a} - \frac{b'}{b}\right]$$

Where

$$c = \sum_{k=0}^{z} U_{z_{jik}} \psi_{ik}$$

$$c' = U_{z_{jik}} \sum_{k=0}^{z} \tau_{ik}$$

$$a = \exp(z\Delta) + \exp((M-z)\Delta)$$

$$a' = \frac{z\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp(z\Delta)\right] + (M-z)\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp((M-z)\Delta)\right]}{\Delta}$$

$$b = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w} U_{w_{jik}} \psi_{ik}\right)\left(\exp(w\Delta) + \exp((M-w)\Delta)\right)\right)$$

and

$$b' = \sum_{w=0}^{C}\left(\begin{array}{l}\left(\exp\sum_{k=0}^{w} U_{w_{jik}} \psi_{ik}\right)\left(\frac{w\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp(w\Delta)\right] + (M-w)\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp((M-w)\Delta)\right]}{\Delta}\right) \\ +\left(\exp(w\Delta) + \exp((M-w)\Delta)\right)\left(U_{w_{jik}} \sum_{k=0}^{w} \tau_{ik}\right)\left(\exp\sum_{k=0}^{w} U_{w_{jik}} \psi_{ik}\right)\end{array}\right).$$

The first partial derivative of the complete data log likelihood with respect to τ_{ik} is:

$$\frac{\partial \ln(L)}{\partial \tau_{ik}} = \sum_{j=1}^{J} \sum_{i=1}^{I} \left[c' - \frac{b'}{b} \right]$$

Where

$$c = \sum_{k=0}^{z} U_{z_{jik}} \psi_{ik}$$

$$c' = U_{z_{jik}} \sum_{d=1}^{D} \alpha_{id}$$

$$b = \sum_{w=0}^{C} \left(\left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp(w\Delta) + \exp((M-w)\Delta) \right) \right)$$

and

$$b' = \sum_{w=1}^{C} \left(\left(U_{w_{jik}} \sum_{d=1}^{D} \alpha_{id} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp(w\Delta) + \exp((M-w)\Delta) \right) \right) .$$

The second partial derivative of the complete data log likelihood with respect to δ_{id} is:

$$\frac{\partial^2 \ln(L)}{\partial \delta_{id}^2} = \sum_{j=1}^{J}\sum_{i=1}^{I}\left[\left(\frac{fe' - ef'}{f^2}\right) - \left(\frac{hg' - gh'}{h^2}\right)\right]$$

Where

$$e = z\alpha_{id}^2(\delta_{id} - \theta_{jd})\left[\exp(z\Delta)\right] + (M - z)\alpha_{id}^2(\delta_{id} - \theta_{jd})\left[\exp((M - z)\Delta)\right]$$

$$e' = \left(z\alpha_{id}^2\left[\exp(z\Delta)\right] + (M - z)\alpha_{id}^2\left[\exp((M - z)\Delta)\right]\right) + \left(\frac{\left(z\alpha_{id}^2(\delta_{id} - \theta_{jd})\right)^2\left[\exp(z\Delta)\right] + \left((M - z)\alpha_{id}^2(\delta_{id} - \theta_{jd})\right)^2\left[\exp((M - z)\Delta)\right]}{\Delta}\right)$$

$$f = \left(\exp[z\Delta] + \exp[(M - z)\Delta]\right)\Delta$$

$$f' = a_\delta\left(\alpha_{id}^2(\delta_{id} - \theta_{jd})/\Delta\right) + \Delta a'_\delta$$

$$g = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(w\alpha_{id}^2(\delta_{id} - \theta_{jd})\left[\exp(w\Delta)\right] + (M - w)\alpha_{id}^2(\delta_{id} - \theta_{jd})\left[\exp((M - w)\Delta)\right]\right)\right)$$

$$g' = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\left(w\alpha_{id}^2\left[\exp(w\Delta)\right] + (M - w)\alpha_{id}^2\left[\exp((M - w)\Delta)\right]\right) + \left(\frac{\left(w\alpha_{id}^2(\delta_{id} - \theta_{jd})\right)^2\left[\exp(w\Delta)\right] + \left((M - w)\alpha_{id}^2(\delta_{id} - \theta_{jd})\right)^2\left[\exp((M - w)\Delta)\right]}{\Delta}\right)\right)\right)$$

$$h = \left(\sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp[w\Delta] + \exp[(M - w)\Delta]\right)\right)\right)\Delta$$

and

$$h' = b_\delta\left(\alpha_{id}^2(\delta_{id} - \theta_{jd})/\Delta\right) + \Delta b'_\delta \quad .$$

The second partial derivative of the complete data log likelihood with respect to α_{id} is:

$$\frac{\partial^2 \ln(L)}{\partial \alpha_{id}^2} = \sum_{j=1}^{J}\sum_{i=1}^{I}\left[\left(\frac{fe' - ef'}{f^2}\right) - \left(\frac{hg' - gh'}{h^2}\right)\right]$$

Where

$$e = z\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp(z\Delta)\right] + (M - z)\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp((M - z)\Delta)\right]$$

$$e' = \left(z(\theta_{jd} - \delta_{id})^2\left[\exp(z\Delta)\right] + (M - z)(\theta_{jd} - \delta_{id})^2\left[\exp((M - z)\Delta)\right]\right) + \left(\frac{\left(z\alpha_{id}(\theta_{jd} - \delta_{id})^2\right)^2\left[\exp(z\Delta)\right] + \left((M - z)\alpha_{id}(\theta_{jd} - \delta_{id})^2\right)^2\left[\exp((M - z)\Delta)\right]}{\Delta}\right)$$

$$f = \left(\exp[z\Delta] + \exp[(M - z)\Delta]\right)\Delta$$

$$f' = a_\alpha\left(\alpha_{id}(\theta_{jd} - \delta_{id})^2/\Delta\right) + \Delta a'_\alpha$$

$$g = \sum_{w=0}^{C}\left(\begin{array}{l}\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\left(w\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp(w\Delta)\right] + (M - w)\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp((M - w)\Delta)\right]\right)\right) \\ + \left(\left(\exp[w\Delta] + \exp[(M - w)\Delta]\right)\left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\Delta\right)\end{array}\right)$$

$$g' = \sum_{w=0}^{C}\left(\begin{array}{l}\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\begin{array}{l}\left(w(\theta_{jd} - \delta_{id})^2\left[\exp(w\Delta)\right] + (M - w)(\theta_{jd} - \delta_{id})^2\left[\exp((M - w)\Delta)\right]\right) + \\ \left(\frac{\left(w\alpha_{id}(\theta_{jd} - \delta_{id})^2\right)^2\left[\exp(w\Delta)\right] + \left((M - w)\alpha_{id}(\theta_{jd} - \delta_{id})^2\right)^2\left[\exp((M - w)\Delta)\right]}{\Delta}\right)\end{array}\right) \\ + \left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(w\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp(w\Delta)\right] + (M - w)\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp((M - w)\Delta)\right]\right) \\ + \left(\exp[w\Delta] + \exp[(M - w)\Delta]\right)\left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\alpha_{id}(\theta_{jd} - \delta_{id})^2/\Delta\right) + \Delta\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)^2 \\ + \left(\left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\Delta\right)\left(\frac{\left(w\alpha_{id}(\theta_{jd} - \delta_{id})^2\right)^2\left[\exp(w\Delta)\right] + \left((M - w)\alpha_{id}(\theta_{jd} - \delta_{id})^2\right)^2\left[\exp((M - w)\Delta)\right]}{\Delta}\right)\end{array}\right)$$

$$h = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp[w\Delta] + \exp[(M - w)\Delta]\right)\right)\Delta$$

and

$$h' = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp(w\Delta) + \exp((M - w)\Delta)\right)\right)\left(\alpha_{id}(\theta_{jd} - \delta_{id})^2/\Delta\right) + \sum_{w=0}^{C}\left(\begin{array}{l}\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\frac{w\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp(w\Delta)\right] + (M - w)\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp((M - w)\Delta)\right]}{\Delta}\right) \\ + \left(\exp(w\Delta) + \exp((M - w)\Delta)\right)\left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\end{array}\right)\Delta .$$

The second partial derivative of the complete data log likelihood with respect to τ_{ik} is:

$$\frac{\partial^2 \ln(L)}{\partial \tau_{ik}^2} = \sum_{j=1}^{J}\sum_{i=1}^{I}\left[-\left(\frac{hg' - gh'}{h^2}\right)\right]$$

Where

$$g = \sum_{w=1}^{C}\left(\left(U_{w_{jik}}\sum_{d=1}^{D}\alpha_{id}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp(w\Delta)+\exp((M-w)\Delta)\right)\right)$$

$$g' = \sum_{w=1}^{C}\left(\left(U_{w_{jik}}\sum_{d=1}^{D}\alpha_{id}\right)^2\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp(w\Delta)+\exp((M-w)\Delta)\right)\right)$$

$$h = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp(w\Delta)+\exp((M-w)\Delta)\right)\right)$$

$$h' = \sum_{w=0}^{C}\left(\left(U_{w_{jik}}\sum_{d=1}^{D}\alpha_{id}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp(w\Delta)+\exp((M-w)\Delta)\right)\right)$$

The second mixed partial derivative of the complete data log likelihood with respect to δ_{id} and δ_{id*} is:

$$\frac{\partial^2 \ln(L)}{\partial\delta_{id}\delta_{id*}} = \sum_{j=1}^{J}\sum_{i=1}^{I}\left[\left(\frac{fe' - ef'}{f^2}\right) - \left(\frac{hg' - gh'}{h^2}\right)\right]$$

Where

$$e = z\alpha_{id}^2(\delta_{id} - \theta_{jd})\left[\exp(z\Delta)\right] + (M - z)\alpha_{id}^2(\delta_{id} - \theta_{jd})\left[\exp((M - z)\Delta)\right]$$

$$e' = \frac{z^2\alpha_{id}^2(\delta_{id} - \theta_{jd})\alpha_{id*}^2(\delta_{id*} - \theta_{jd*})\left[\exp(z\Delta)\right] + (M - z)^2\alpha_{id}^2(\delta_{id} - \theta_{jd})\alpha_{id*}^2(\delta_{id*} - \theta_{jd*})\left[\exp((M - z)\Delta)\right]}{\Delta}$$

$$f = \left(\exp[z\Delta] + \exp[(M - z)\Delta]\right)\Delta$$

$$f' = a_\delta\left(\alpha_{id*}^2(\delta_{id*} - \theta_{jd*})/\Delta\right) + \Delta a'_{\delta*}$$

$$g = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(w\alpha_{id}^2(\delta_{id} - \theta_{jd})\left[\exp(w\Delta)\right] + (M - w)\alpha_{id}^2(\delta_{id} - \theta_{jd})\left[\exp((M - w)\Delta)\right]\right)\right)$$

$$g' = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\frac{w^2\alpha_{id}^2(\delta_{id} - \theta_{jd})\alpha_{id*}^2(\delta_{id*} - \theta_{jd*})\left[\exp(w\Delta)\right] + (M - w)^2\alpha_{id}^2(\delta_{id} - \theta_{jd})\alpha_{id*}^2(\delta_{id*} - \theta_{jd*})\left[\exp((M - w)\Delta)\right]}{\Delta}\right)\right)$$

$$h = \left(\sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp[w\Delta] + \exp[(M - w)\Delta]\right)\right)\right)\Delta$$

$$h' = b_\delta\left(\alpha_{id*}^2(\delta_{id*} - \theta_{jd*})/\Delta\right) + \Delta b'_{\delta*} \ .$$

The second mixed partial derivative of the complete data log likelihood with respect to α_{id} and α_{id*} is:

$$\frac{\partial^2 \ln(L)}{\partial \alpha_{id}\alpha_{id*}} = \sum_{j=1}^{J}\sum_{i=1}^{I}\left[\left(\frac{fe' - ef'}{f^2}\right) - \left(\frac{hg' - gh'}{h^2}\right)\right]$$

Where

$$e = z\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp(z\Delta)\right] + (M - z)\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp((M - z)\Delta)\right]$$

$$e' = \frac{z^2\alpha_{id}(\theta_{jd} - \delta_{id})^2\alpha_{id*}(\theta_{jd*} - \delta_{id*})^2\left[\exp(z\Delta)\right] + (M - z)^2\alpha_{id}(\theta_{jd} - \delta_{id})^2\alpha_{id*}(\theta_{jd*} - \delta_{id*})^2\left[\exp((M - z)\Delta)\right]}{\Delta}$$

$$f = \left(\exp[z\Delta] + \exp[(M - z)\Delta]\right)\Delta$$

$$f' = a_\alpha\left(\alpha_{id*}(\theta_{jd*} - \delta_{id*})^2/\Delta\right) + \Delta a'_{\alpha*}$$

$$g = \sum_{w=0}^{C}\left(\begin{array}{l}\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\left(w\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp(w\Delta)\right] + (M - w)\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp((M - w)\Delta)\right]\right)\right) \\ + \left(\left(\exp[w\Delta] + \exp[(M - w)\Delta]\right)\left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\Delta\right)\end{array}\right)$$

$$g' = \sum_{w=0}^{C}\left(\begin{array}{l}\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\frac{w^2\alpha_{id}(\theta_{jd} - \delta_{id})^2\alpha_{id*}(\theta_{jd*} - \delta_{id*})^2\left[\exp(w\Delta)\right] + (M - w)^2\alpha_{id}(\theta_{jd} - \delta_{id})^2\alpha_{id*}(\theta_{jd*} - \delta_{id*})^2\left[\exp((M - w)\Delta)\right]}{\Delta}\right) \\ \quad + \left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(w\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp(w\Delta)\right] + (M - w)\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp((M - w)\Delta)\right]\right) \\ \quad + \left(\exp[w\Delta] + \exp[(M - w)\Delta]\right)\left(\left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\alpha_{id*}(\theta_{jd*} - \delta_{id*})^2/\Delta\right) + \left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)^2\Delta\right) \\ \quad + \left(\left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\Delta\right)\left(\frac{w\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp(w\Delta)\right] + (M - w)\alpha_{id}(\theta_{jd} - \delta_{id})^2\left[\exp((M - w)\Delta)\right]}{\Delta}\right)\end{array}\right)$$

$$h = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp[w\Delta] + \exp[(M - w)\Delta]\right)\right)\Delta$$

and

$$h' = \sum_{w=0}^{C}\left(\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\exp(w\Delta) + \exp((M - w)\Delta)\right)\right)\left(\alpha_{id*}(\theta_{jd*} - \delta_{id*})^2/\Delta\right)$$
$$+ \sum_{w=0}^{C}\left(\begin{array}{l}\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\left(\frac{w\alpha_{id*}(\theta_{jd*} - \delta_{id*})^2\left[\exp(w\Delta)\right] + (M - w)\alpha_{id*}(\theta_{jd*} - \delta_{id*})^2\left[\exp((M - w)\Delta)\right]}{\Delta}\right) \\ + \left(\exp(w\Delta) + \exp((M - w)\Delta)\right)\left(U_{w_{jik}}\sum_{k=0}^{w}\tau_{ik}\right)\left(\exp\sum_{k=0}^{w}U_{w_{jik}}\psi_{ik}\right)\end{array}\right)\Delta\ .$$

The second mixed partial derivative of the complete data log likelihood with respect to τ_{ik} and τ_{ik*} is:

$$\frac{\partial^2 \ln(L)}{\partial \tau_{ik} \tau_{ik*}} = \sum_{j=1}^{J} \sum_{i=1}^{I} \left[-\left(\frac{hg' - gh'}{h^2} \right) \right]$$

Where

$$g = \sum_{w=1}^{C} \left(\left(U_{w_{jik}} \sum_{d=1}^{D} \alpha_{id} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp(w\Delta) + \exp((M-w)\Delta) \right) \right)$$

$$g' = \sum_{w=1}^{C} \left(\left(U_{w_{jik}} \sum_{d=1}^{D} \alpha_{id} \right) \left(U_{w_{jik*}} \sum_{d=1}^{D} \alpha_{id} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp(w\Delta) + \exp((M-w)\Delta) \right) \right)$$

$$h = \sum_{w=0}^{C} \left(\left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp(w\Delta) + \exp((M-w)\Delta) \right) \right)$$

$$h' = \sum_{w=0}^{C} \left(\left(U_{w_{jik}} \sum_{d=1}^{D} \alpha_{id} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp(w\Delta) + \exp((M-w)\Delta) \right) \right) .$$

The second mixed partial derivative of the complete data log likelihood with respect to δ_{id} and α_{id} is:

$$\frac{\partial^2 \ln(L)}{\partial \delta_{id} \alpha_{id}} = \sum_{j=1}^{J} \sum_{i=1}^{I} \left[c' + \left(\frac{fe' - ef'}{f^2} \right) - \left(\frac{hg' - gh'}{h^2} \right) \right]$$

Where

$$c = \sum_{k=0}^{z} U_{z_{jik}} \psi_{ik}$$

$$c' = U_{z_{jik}} \sum_{k=0}^{z} \tau_{ik}$$

$$e = z\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp(z\Delta) \right] + (M - z)\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp((M - z)\Delta) \right]$$

$$e' = \left(\frac{z\alpha_{id}^2 (\delta_{id} - \theta_{jd}) z\alpha_{id} (\theta_{jd} - \delta_{id})^2 \left[\exp(z\Delta) \right] + (M - z)\alpha_{id}^2 (\delta_{id} - \theta_{jd})(M - z)\alpha_{id} (\theta_{jd} - \delta_{id})^2 \left[\exp((M - z)\Delta) \right]}{\Delta} \right)$$
$$+ \left(2z\alpha_{id} (\delta_{id} - \theta_{jd}) \left[\exp(z\Delta) \right] + 2(M - z)\alpha_{id} (\delta_{id} - \theta_{jd}) \left[\exp((M - z)\Delta) \right] \right)$$

$$f = \left(\exp[z\Delta] + \exp[(M - z)\Delta] \right) \Delta$$

$$f' = a_\alpha \left(\alpha_{id} (\theta_{jd} - \delta_{id})^2 / \Delta \right) + \Delta a'_\alpha$$

$$g = \sum_{w=0}^{C} \left(\left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(w\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp(w\Delta) \right] + (M - w)\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp((M - w)\Delta) \right] \right) \right)$$

$$g' = \sum_{w=0}^{C} \left(\begin{array}{l} \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\left(\frac{w\alpha_{id}^2 (\delta_{id} - \theta_{jd}) w\alpha_{id} (\theta_{jd} - \delta_{id})^2 \left[\exp(w\Delta) \right] + (M - w)\alpha_{id}^2 (\delta_{id} - \theta_{jd})(M - w)\alpha_{id} (\theta_{jd} - \delta_{id})^2 \left[\exp((M - w)\Delta) \right]}{\Delta} \right) \right. \\ \qquad \left. + \left(2w\alpha_{id} (\delta_{id} - \theta_{jd}) \left[\exp(w\Delta) \right] + 2(M - w)\alpha_{id} (\delta_{id} - \theta_{jd}) \left[\exp((M - w)\Delta) \right] \right) \right) \\ + \left(U_{w_{jik}} \sum_{k=0}^{w} \tau_{ik} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(w\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp(w\Delta) \right] + (M - w)\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp((M - w)\Delta) \right] \right) \end{array} \right)$$

$$h = \left(\sum_{w=0}^{C} \left(\left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp[w\Delta] + \exp[(M - w)\Delta] \right) \right) \right) \Delta$$

$$h' = b_\alpha \left(\alpha_{id} (\theta_{jd} - \delta_{id})^2 / \Delta \right) + \Delta b'_\alpha \ .$$

The second mixed partial derivative of the complete data log likelihood with respect to δ_{id} and α_{id*} is:

$$\frac{\partial^2 \ln(L)}{\partial \delta_{id} \alpha_{id*}} = \sum_{j=1}^{J} \sum_{i=1}^{I} \left[c' + \left(\frac{fe' - ef'}{f^2} \right) - \left(\frac{hg' - gh'}{h^2} \right) \right]$$

Where

$$c = \sum_{k=0}^{z} U_{z_{jik}} \psi_{ik}$$

$$c' = U_{z_{jik}} \sum_{k=0}^{z} \tau_{ik}$$

$$e = z\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp(z\Delta) \right] + (M - z)\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp((M - z)\Delta) \right]$$

$$e' = \frac{z\alpha_{id}^2 (\delta_{id} - \theta_{jd}) z\alpha_{id*} (\theta_{jd*} - \delta_{id*})^2 \left[\exp(z\Delta) \right] + (M - z)\alpha_{id}^2 (\delta_{id} - \theta_{jd})(M - z)\alpha_{id*} (\theta_{jd*} - \delta_{id*})^2 \left[\exp((M - z)\Delta) \right]}{\Delta}$$

$$f = \left(\exp[z\Delta] + \exp[(M - z)\Delta] \right) \Delta$$

$$f' = a_\alpha \left(\alpha_{id*} (\theta_{jd*} - \delta_{id*})^2 / \Delta \right) + \Delta a'_{\alpha*}$$

$$g = \sum_{w=0}^{C} \left(\left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(w\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp(w\Delta) \right] + (M - w)\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp((M - w)\Delta) \right] \right) \right)$$

$$g' = \sum_{w=0}^{C} \left(\begin{array}{l} \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\frac{w\alpha_{id}^2 (\delta_{id} - \theta_{jd}) w\alpha_{id*} (\theta_{jd*} - \delta_{id*})^2 \left[\exp(w\Delta) \right] + (M - w)\alpha_{id}^2 (\delta_{id} - \theta_{jd})(M - w)\alpha_{id*} (\theta_{jd*} - \delta_{id*})^2 \left[\exp((M - w)\Delta) \right]}{\Delta} \right) \\ + \left(U_{w_{jik}} \sum_{k=0}^{w} \tau_{ik} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(w\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp(w\Delta) \right] + (M - w)\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp((M - w)\Delta) \right] \right) \end{array} \right)$$

$$h = \left(\sum_{w=0}^{C} \left(\left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp[w\Delta] + \exp[(M - w)\Delta] \right) \right) \right) \Delta$$

$$h' = b_\alpha \left(\alpha_{id*} (\theta_{jd*} - \delta_{id*})^2 / \Delta \right) + \Delta b'_{\alpha*} \ .$$

The second mixed partial derivative of the complete data log likelihood with respect to δ_{id} and τ_{ik} is:

$$\frac{\partial^2 \ln(L)}{\partial \delta_{id} \tau_{ik}} = \sum_{j=1}^{J} \sum_{i=1}^{I} \left[-\left(\frac{hg' - gh'}{h^2} \right) \right]$$

Where

$$g = \sum_{w=0}^{C} \left(\left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(w\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp(w\Delta) \right] + (M-w)\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp((M-w)\Delta) \right] \right) \right)$$

$$g' = \sum_{w=0}^{C} \left(\left(U_{w_{jik*}} \sum_{d=1}^{D} \alpha_{id} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(w\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp(w\Delta) \right] + (M-w)\alpha_{id}^2 (\delta_{id} - \theta_{jd}) \left[\exp((M-w)\Delta) \right] \right) \right)$$

$$h = \left(\sum_{w=0}^{C} \left(\left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp[w\Delta] + \exp[(M-w)\Delta] \right) \right) \right) \Delta$$

and

$$h' = \left(\sum_{w=0}^{C} \left(\left(U_{w_{jik*}} \sum_{d=1}^{D} \alpha_{id} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp[w\Delta] + \exp[(M-w)\Delta] \right) \right) \right) \Delta \ .$$

The second mixed partial derivative of the complete data log likelihood with respect to α_{id} and τ_{ik} is:

$$\frac{\partial^2 \ln(L)}{\partial \alpha_{id} \tau_{ik}} = \sum_{j=1}^{J} \sum_{i=1}^{I} \left[l' - \left(\frac{hg' - gh'}{h^2} \right) \right]$$

Where

$$l = U_{z_{jik}} \sum_{k=0}^{z} \tau_{ik}$$

$$l' = U_{z_{jik}}$$

$$g = \sum_{w=0}^{C} \left(\begin{array}{l} \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\left(w\alpha_{id} (\theta_{jd} - \delta_{id})^2 \left[\exp(w\Delta) \right] + (M-w)\alpha_{id} (\theta_{jd} - \delta_{id})^2 \left[\exp((M-w)\Delta) \right] \right) \right) \\ + \left(\left(\exp[w\Delta] + \exp[(M-w)\Delta] \right) \left(U_{w_{jik}} \sum_{k=0}^{w} \tau_{ik} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \Delta \right) \end{array} \right)$$

$$g' = \sum_{w=0}^{C} \left(\begin{array}{l} \left(U_{w_{jik^*}} \sum_{d=1}^{D} \alpha_{id} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\left(w\alpha_{id} (\theta_{jd} - \delta_{id})^2 \left[\exp(w\Delta) \right] + (M-w)\alpha_{id} (\theta_{jd} - \delta_{id})^2 \left[\exp((M-w)\Delta) \right] \right) \right) \\ + \left(\left(\exp[w\Delta] + \exp[(M-w)\Delta] \right) \left(\left(U_{w_{jik}} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) + \left(U_{w_{jik}} \sum_{k=0}^{w} \tau_{ik} \right) \left(U_{w_{jik^*}} \sum_{d=1}^{D} \alpha_{id} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \right) \Delta \right) \end{array} \right)$$

$$h = \sum_{w=0}^{C} \left(\left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp[w\Delta] + \exp[(M-w)\Delta] \right) \right) \Delta$$

$$h' = \sum_{w=0}^{C} \left(\left(U_{w_{jik^*}} \sum_{d=1}^{D} \alpha_{id} \right) \left(\exp \sum_{k=0}^{w} U_{w_{jik}} \psi_{ik} \right) \left(\exp[w\Delta] + \exp[(M-w)\Delta] \right) \right) \Delta \ .$$

REFERENCES

Andrich, D. (1988). The application of an unfolding model of the PIRT type for the measurement of attitude. *Applied Psychological Measurement, 12*, 33-51.

Andrich, D. (1989). A probabilistic IRT model for unfolding preference data. *Applied Psychological Measurement, 13*, 193-216.

Andrich, D., & Luo, G. (1993). A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses. *Applied Psychological Measurement, 17*(3), 253-276.

Bellman, R. (1957). *Dynamic Programming*. Princeton: New Jersey.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-472). Reading, MA: Addison Wesley.

Bock R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika, 46,* 443-459.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP Estimation of ability in a microcomputer environment. *Applied Psychological Measurement, 6,* 431–444.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing, 16*(5), 1190-1208.

Cai, L. (2010a). A two-tier full-information item factor analysis model with applications. *Psychometrika, 75*(4), 581-612.

Cai, L. (2010b). High-dimensional exploratory item factor analysis by a Metropolis–Hastings Robbins–Monro algorithm. *Psychometrika, 75*(1), 33-57.

Cai, L. (2010c). Metropolis-Hastings Robbins-Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics, 35*(3), 307-335.

Carroll, J. D., & Chang, J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*, 283-319.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1-29. URL: http://www.jstatsoft.org/v48/i06/.

Chalmers, R. P., & Flora, D. B. (2014). Maximum-likelihood estimation of noncompensatory IRT models with the MH-RM algorithm. *Applied Psychological Measurement*, 1-20.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician, 49*(4), 327-335.

Davison, M. L. (1977). On a metric, unidimensional unfolding model for attitudinal and developmental data. *Psychometrika, 42*, 523-548.

De la Torre, J., Stark, S., & Chernyshenko, O. S. (2006). Markov Chain Monte Carlo estimation of item parameters for the generalized graded unfolding model. *Applied Psychological Measurement, 30*, 216-232.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression, Second Edition*. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Gelfand, A. E., Dey, D. K., & Chang, H. (1992). *Model determination using predictive distributions with implementation via sampling-based methods* (No. TR-462). Stanford University, CA, Department of Statistics.

Geisser, S., & Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association, 74*, 153–160.

Hastings, W. K. (1970). Monte Carlo simulation methods using Markov chains and their applications. *Biometrika, 57*, 97-109.

Hill, M.O., & Gauch, H.G. (1980). Detrended correspondence analysis: An improved ordination technique. *Vegetatio, 42*, 47-58.

Javaris, K.N., & Ripley, B.D. (2007). An "unfolding" latent variable model for Likert attitude data: Drawing inferences adjusted for response style. *Journal of the American Statistical Association, 102*, 454-463.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, *10*(4), 325-337.

Meng, X.L. & Schilling, S.G. (1996). Fitting full-information factor models and an empirical investigation of bridge sampling. *Journal of the American Statistical Association, 91*, 1254–1267.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087-1092.

Mislevy, R.J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51*(2), 177-195.

Muraki, E., & Carlson, J. E. (1995). Full-information factor analysis for polytomous item responses. *Applied Psychological Measurement, 19*, 73-90.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. A. S. S. (2007). The vegan package. *Community ecology package*, 631-637.

Osgood, C.E., Succi, G., & Tannenbaum, P. (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.

Paek, I., & Cai, L. (2014). A comparison of item parameter standard error estimation procedures for unidimensional and multidimensional item response theory modeling. *Educational and Psychological Measurement*, *74*(1), 58-76.

Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, *24*(2), 146-178.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer-Verlag.

Reckase, M. D., & McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Applied psychological measurement*, *15*(4), 361-373.

Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, *22*, 400-407.

Roberts, J. S. (1995). Item response theory approaches to attitude measurement. (Doctoral dissertation, University of South Carolina, Columbia, 1995). *Dissertation Abstracts International, 56*, 7089B.

Roberts, J.S., Donoghue, J.R., & Laughlin, J.E. (2000). A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*(1), 3-32.

Roberts, J.S., & Laughlin, J.E. (1996). A unidimensional item response model for unfolding responses from a graded disagree-agree scale. *Applied Psychological Measurement, 20,* 231-255.

Roberts, J.S., Laughlin, J.E., & Wedell, D.H. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Educational and Psychological Measurement, 59,* 211-233.

Roberts, J.S., & Shim, H. (July, 2010). *Multidimensional unfolding with item response theory: The multidimensional generalized graded unfolding model*. Paper presented at the 2010 Annual Meeting of the Psychometric Society, Athens, Georgia.

Roberts, J. S., & Sparks, J. (2015, April). *Mapping the emotion space with the MGGUM*. Poster session presented at the meeting of the National Council on Measurement in Education, Chicago, IL.

Roberts, J.S., & Thompson, V.M. (2011). Marginal maximum a posteriori item parameter estimation for the generalized graded unfolding model. *Applied Psychological Measurement*.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 116-1178.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of research in Personality, 11*(3), 273-294.

Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika, 70*(3), 533-555.

Schlosberg, H. (1941). A scale for the judgment of facial expressions. *Journal of Experimental Psychology, 29*, 497-510.

Schlosberg, H. (1952). The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology, 44*(4), 229-237.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–64.

Sparks, J. (2015). *A hybrid external multidimensional unfolding approach for the analysis of self-similarity emotion judgments.* Unpublished master's thesis, Georgia Institute of Technology, Atlanta, Georgia.

Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology, 91*(1).

Sympson, J. B. (1978). A model for testing with multidimensional items. In *Proceedings of the 1977 computerized adaptive testing conference* (pp. 82-98). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.

Tay, L., Drasgow, F., Rounds, J., & Williams, B. A. (2009). Fitting measurement models to vocational interest data: Are dominance models ideal? *Journal of Applied Psychology, 94*(5), 1287.

Thompson, V. M. (2014). *Marginal Bayesian parameter estimation in the multidimensional generalized graded unfolding model*. Doctoral dissertation. Georgia Institute of Technology.

Thurstone, L.L. (1928). Attitudes can be measured. *The American Journal of Sociology 33*, 529-553.

Wang, W., & Wu, S. (2015). Confirmatory multidimensional IRT unfolding models for graded-response items. *Applied Psychological Measurement*.

Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika, 45,* 479-494.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.